

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2003, 11:57:49 ; Search time 10.9842 Seconds
(without alignments)
4722.097 Million cell updates/sec

Title: US-09-970-318-2
Perfect score: 1092
Sequence: 1 FLAQRVCIPNDAGTADNRV.....YTTETAANLHRYAYDREIT 201

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1092	100.0	1546	4	O75445
2	1086	99.5	1546	4	O75445
3	759	69.5	1461	11	Q9JLP3
4	728	66.7	1512	11	Q8K3K1
5	294	26.9	214	11	Q9D128
6	186	17.0	569	13	O57339
7	185	16.9	603	13	O42140
8	181	16.6	604	11	Q92429
9	175	16.0	670	5	O95YK2
10	173.5	15.9	3695	4	Q8TDF8
11	163.5	15.0	600	5	O9NL27
12	163.5	15.0	650	5	O9NL28
13	161	14.7	3704	5	P91904
14	160.5	14.7	602	13	O42203
15	160	14.7	529	4	Q8N2D6
16	159.5	14.6	555	5	Q9NFW6

17	159.5	14.6	1785	13	O8JHV7
18	149	13.6	1593	13	O8JHV8
19	142.5	13.0	1792	13	O57484
20	141.5	13.0	1486	4	O14637
21	140.5	12.9	667	5	O9VY25
22	139	12.7	628	11	O9J133
23	138	12.6	610	5	O96659
24	138	12.6	3712	5	O9VRW0
25	136.5	12.5	628	4	O9HB63
26	136.5	12.5	1623	5	O9U3U7
27	136.5	12.5	1808	5	O44565
28	135.5	12.4	628	4	O9B2P1
29	135	12.4	1827	13	O8JHV6
30	133.5	12.2	3102	5	O45614
31	131	12.0	168	13	O91435
32	131	12.0	761	4	O9UHI2
33	131	12.0	1631	4	O9Y6U6
34	130.5	12.0	1799	11	O8ROY0
35	117.5	10.8	695	11	O8C9J2
36	116.5	10.7	530	11	O8R4F2
37	116.5	10.7	530	11	O8VIP8
38	116.5	10.7	564	11	O8VIP6
39	116.5	10.7	589	11	O8R4F1
40	116.5	10.7	589	11	O8VIP7
41	115.5	10.6	530	4	O96CW9
42	115.5	10.6	549	4	O96JH0
43	103.5	9.5	364	4	O9I212
44	103.5	9.5	438	4	O8N633
45	102	9.3	580	11	O9QY49

ALIGNMENTS

RESULT 1

O75445	PRELIMINARY;	PRT; 1546 AA.
ID	O75445	
AC	O75445;	
DT	01-NOV-1998 (Tremblrel. 08, Created)	
DT	01-JUN-2002 (Tremblrel. 21, Last sequence update)	
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)	
DE	Usher syndrome type IIA protein.	
GN	USH2A.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=98288356; PubMed=9624053;	
RA	Eudy J.D., Weston M.D., Yao S.F., Hoover D.M., Rehm H.L., Ahmad I.,	
RA	Ma-Edmonds M., Yan D., Cheng J.J., Beisel K.W., Ayuso C., Cremers C.,	
RA	Davenport S., Moller C., Talmadge C.B., Tamayo M., Swaroop A.,	
RA	Morton C.C., Kimberling W.J., Sumegi J.;	
RT	"Mutation of a gene encoding a protein with extracellular matrix	
RT	motifs in Usher syndrome type IIA.";	
RL	Science 280:1753-1757(1998).	
RL	[2]	
RN	SEQUENCE FROM N.A.	
RP	Eudy J.D., Yao S.F., Cheng J.J., Weston M.D., Sumegi J.;	
RA	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.	
RL	EMBL; AF055580; AAC23748.2; -	
DR	HSSP; P02468; IKLO.	
DR	Genew; HGNC:12601; USH2A.	
DR	InterPro; IPR006209; EGF like.	
DR	InterPro; IPR003961; FN III.	
DR	InterPro; IPR006558; LamG like.	
DR	InterPro; IPR002049; Laminin_EGF.	
DR	InterPro; IPR001886; LamNT.	
DR	Pfam; PF00041; fn3; 4.	
DR	Pfam; PF00053; laminin_EGF; 9.	
DR	Pfam; PF00055; laminin_Nterm; 1.	
DR	PRINTS; PR00011; EGF-LAMININ.	

```

DR SMART; SM00180; EGF Lam; 10.
DR SMART; SM00060; FN3; 4.
DR SMART; SM00560; LamGL; 1.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF 1; 7.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
KW Laminin EGF-like domain.
SQ SEQUENCE 1546 AA; 170985 MW; 49CD0A95A614959F CRC64;

Query Match 100.0%; Score 1092; DB 4; Length 1546;
Best Local Similarity 100.0%; Pred. No. 1.8e-97;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLAQRVCIPNDAGDTADNRVSRNLNPEAHPLSPVNDNDVGTSWSNVFTNITQLNQGVTTIS 60
DB 313 PLAQRVCIPNDAGDTADNRVSRNLNPEAHPLSPVNDNDVGTSWSNVFTNITQLNQGVTTIS 372
QY 61 VDLENGQYQVFIILIOFFSPQPTTEIRIQKKNLSLWEDWQYFARNCGAFGMKNNGDLEK 120
DB 373 VDLENGQYQVFIILIOFFSPQPTTEIRIQKKNLSLWEDWQYFARNCGAFGMKNNGDLEK 432
QY 121 PDSVNCQLQSNFTPYSGNGVTFSLTPGPNRYRFGYNNFYNTPSLOBSVKATQIRPHFHQ 180
DB 433 PDSVNCQLQSNFTPYSGNGVTFSLTPGPNRYRFGYNNFYNTPSLOBSVKATQIRPHFHQ 492
QY 181 YTTTAVNLHRHYAVDEIT 201
DB 493 YTTTAVNLHRHYAVDEIT 513

RESULT 2
Q9NS27 Q9NS27 PRELIMINARY; PRT; 1546 AA.
ID Q9NS27;
AC Q9NS27;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Usher syndrome type IIA protein.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20206315; PubMed=10729113;
RA Weston M.D., Eudy J.D., Fugita S., Yao S.-F., Usami S., Cremers C.,
RA Greenburg J., Ramessar R., Martini A., Moller C., Smith R.J.,
RA Sumegi J., Kimberling W.J.;
RT "Genomic structure and identification of novel mutations in usherin,
RT the gene responsible for Usher syndrome type IIA.";
RL Am. J. Hum. Genet. 66:1199-1210(2000).
DR EMBL; AF091889; AAF75819.1; JOINED.
DR EMBL; AF091873; AAF75819.1; JOINED.
DR EMBL; AF091874; AAF75819.1; JOINED.
DR EMBL; AF091875; AAF75819.1; JOINED.
DR EMBL; AF091876; AAF75819.1; JOINED.
DR EMBL; AF091877; AAF75819.1; JOINED.
DR EMBL; AF091878; AAF75819.1; JOINED.
DR EMBL; AF091879; AAF75819.1; JOINED.
DR EMBL; AF091880; AAF75819.1; JOINED.
DR EMBL; AF091881; AAF75819.1; JOINED.
DR EMBL; AF091882; AAF75819.1; JOINED.
DR EMBL; AF091883; AAF75819.1; JOINED.
DR EMBL; AF091884; AAF75819.1; JOINED.
DR EMBL; AF091885; AAF75819.1; JOINED.
DR EMBL; AF091886; AAF75819.1; JOINED.
DR EMBL; AF091887; AAF75819.1; JOINED.
DR EMBL; AF091888; AAF75819.1; JOINED.
DR HSP; P02468; IUKO.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR006558; LamG like.
DR InterPro; IPR002049; Laminin_EGF.

```

```

DR InterPro; IPR001886; LamNT.
DR Pfam; PF00041; fn3; 4.
DR Pfam; PF00053; laminin_EGF; 10.
DR Pfam; PF00055; laminin_Nterm; 1.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00180; EGF_Lam; 10.
DR SMART; SM00060; FN3; 4.
DR SMART; SM00560; LamGL; 1.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF 1; 7.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
KW Laminin EGF-like domain.
SQ SEQUENCE 1546 AA; 171001 MW; 64BD435996C1E17B CRC64;

Query Match 99.5%; Score 1086; DB 4; Length 1546;
Best Local Similarity 99.5%; Pred. No. 6.8e-97;
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PLAQRVCIPNDAGDTADNRVSRNLNPEAHPLSPVNDNDVGTSWSNVFTNITQLNQGVTTIS 60
DB 313 PLAQRVCIPNDAGDTADNRVSRNLNPEAHPLSPVNDNDVGTSWSNVFTNITQLNQGVTTIS 372
QY 61 VDLENGQYQVFIILIOFFSPQPTTEIRIQKKNLSLWEDWQYFARNCGAFGMKNNGDLEK 120
DB 373 VDLENGQYQVFIILIOFFSPQPTTEIRIQKKNLSLWEDWQYFARNCGAFGMKNNGDLEK 432
QY 121 PDSVNCQLQSNFTPYSGNGVTFSLTPGPNRYRFGYNNFYNTPSLOBSVKATQIRPHFHQ 180
DB 433 PDSVNCQLQSNFTPYSGNGVTFSLTPGPNRYRFGYNNFYNTPSLOBSVKATQIRPHFHQ 492
QY 181 YTTTAVNLHRHYAVDEIT 201
DB 493 YTTTAVNLHRHYAVDEIT 513

RESULT 3
Q9JLP3 Q9JLP3 PRELIMINARY; PRT; 1461 AA.
ID Q9JLP3;
AC Q9JLP3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative extracellular matrix protein MUSH2A.
GN MUSH2A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Eudy J.D., Talmadge C.B., Weston M.D., Yao S.-F., Cosgrove D.,
RA Ahmad I., Kimberling W., Sumegi J.;
RT "Isolation and Characterization of the Murine Homolog of the Usher
RT Syndrome Type 2A Gene.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF151717; AAF70550.1; -.
DR HSP; P02468; 1TLE.
DR MGD; MGI:1341292; Ush2a.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR002049; Laminin_EGF.
DR Pfam; PF00041; fn3; 4.
DR Pfam; PF00053; laminin_Nterm; 1.
DR Pfam; PF00055; laminin_EGF; 10.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00180; EGF_Lam; 8.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF 1; 7.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
KW Laminin EGF-like domain; Matrix protein.
SQ SEQUENCE 1461 AA; 160989 MW; 114436D877C38A15 CRC64;

```

Query Match 69.5%; Score 759; DB 11; Length 1461;
 Best Local Similarity 68.5%; Pred. No. 5.4e-65;
 Matches 137; Conservative 25; Mismatches 38; Indels 0; Gaps 0;

QY 1 PLAQRVCIPNDAGTADNRVSRNLPENAPLSPFVNDNDVGTWSVSNVFTNITQLNGQVTIS 60
 DB 310 PSVQVCIPNGAGDTPEHRMSRLNPEAPLSPFINDDDVATSWISHVFTNITQLYEGVAIS 369

QY 61 VDLNGQYQVFIYIIQFPSPQTEIRIQKKNLSLWEDWQYFARNCGAGMKNNGDLEK 120
 DB 370 IDLENGQYQVLTQFSSLPQVAIRIQKKAQSPWEDWQYFARNCSVWGMKNEDLEN 429

QY 121 PDSVNCQLQSNFTPTSGRGNVTFSLTPGPNRYPCYANNFYNTPSLQBSVKATQIRLHFHQ 180
 DB 430 PDSVNCQLQSNFTPTSGRGNVTFSLTPGPNRYPCYANNFYNTPSLQBSVKATQIRLHFHQ 489

QY 181 YTTTAVNLRHRYAYVDEIT 200
 DB 490 YYPAGHTVDWRHRYAYVDEI 509

RESULT 4
 Q8K3K1 PRELIMINARY; PRT; 1512 AA.

ID Q8K3K1
 AC Q8K3K1
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Usherlin.
 GN USH2A.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=22150873; PubMed=12610733;
 RX Huang D., Eudy J.D., Uzvolgyi E., Davis J.R., Talmadge C.B.,
 RA Pretto D., Weston M.D., Lehman J.E., Zhou M., Seemayer T.A., Ahmad I.,
 RA Kimberling W.J., Sumegi J.;
 RT "Identification of the Mouse and Rat Orthologs of the Gene Mutated in
 Usher Syndrome Type IIA and the Cellular Source of USH2A mRNA in
 RT Retina, a Target Tissue of the Disease.";
 RL Genomics 80:195-203(2002).
 DR EMBL; AY077844; AAL78289.1;
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR006558; LamG-like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001896; LamNT.
 DR Pfam; PF00041; fn3; 4.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR SMART; SM00180; EGF_Lam; 10.
 DR SMART; SM00060; FN3; 4.
 DR SMART; SM00560; LamGL; 1.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 6.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 6.
 DR Laminin EGF-like domain.
 KW Laminin
 SQ SEQUENCE 1512 AA; 167591 MW; E37D79F5D00044C1 CRC64;

Query Match 66.7%; Score 728; DB 11; Length 1512;
 Best Local Similarity 66.7%; Pred. No. 6e-62;
 Matches 134; Conservative 22; Mismatches 45; Indels 0; Gaps 0;

QY 1 PLAQRVCIPNDAGTADNRVSRNLPENAPLSPFVNDNDVGTWSVSNVFTNITQLNGQVTIS 60
 DB 309 PSVQVCIPNGVEDTQHRVSRNLPENAPLSPFINDDDVATSWISHVFTNITQLNGQVAIS 368

QY 61 VDLNGQYQVFIYIIQFPSPQTEIRIQKKNLSLWEDWQYFARNCGAGMKNNGDLEK 120
 DB 369 IDLENGQYQVFTIRFSSPQVAIRIQKKAQSLWEDWQYFARNCSVWGMKNNGDLEN 428

QY 121 PDSVNCQLQSNFTPTSGRGNVTFSLTPGPNRYPCYANNFYNTPSLQBSVKATQIRLHFHQ 180
 DB 429 PDSVNCQLQSNFTPTSGRGNVTFSLTPGPNRYPCYANNFYNTPSLQBSVKATQIRLHFHQ 488

QY 181 YTTTAVNLRHRYAYVDEIT 201
 DB 489 FYPAWHTVDSRHYAYVDEIT 509

RESULT 5
 Q9D1Z8 PRELIMINARY; PRT; 214 AA.

ID Q9D1Z8
 AC Q9D1Z8
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE A930011D15Rik protein.
 GN USH2A OR A930011D15RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Retina;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL; AK020845; BAB32226.1; --
 DR MGD; MGI:1341292; Ush2a.
 SQ SEQUENCE 214 AA; 24016 MW; C0040881B1DE5702 CRC64;

Query Match 26.9%; Score 294; DB 11; Length 214;
 Best Local Similarity 74.3%; Pred. No. 1.1e-20;
 Matches 52; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

QY 1 PLAQRVCIPNDAGTADNRVSRNLPENAPLSPFVNDNDVGTWSVSNVFTNITQLNGQVTIS 60
 DB 129 PSVQVCIPNGAGDTPEHRMSRLNPEAPLSPFINDDDVATSWISHVFTNITQLYEGVAIS 188

QY 61 VDLNGQYQV 70
 DB 189 IDLENGQYQV 198

RESULT 6
 O57339 PRELIMINARY; PRT; 569 AA.

ID O57339
 AC O57339
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Netrin-1 (Fragment).
 GN NETRIN-1.

OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA de la Torre J.R., Hoepker V.H., Ming G.-L., Poo M.-m.,
 RA Tessier-Lavigne M., Hemmati-Brivanlou A., Holt C.E.;
 RT "Turning of retinal growth cones in a netrin-1 gradient mediated by
 RT the netrin receptor DCC.";
 RL Neuron 0:0-0(1997).
 DR EMBL; AF033341; AAB87983.1; --
 DR HSSP; P02468; 1TLE.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001886; LamNT.
 DR InterPro; IPR001134; Netrin_C.
 DR Pfam; PF00053; laminin_EGF; 3.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR Pfam; PF01759; NTR; 1.
 DR PRINTS; PR00011; EGFLAMININ.
 DR ProDom; PD002082; LamNT; 1.
 DR SMART; SM00643; C345C; 1.
 DR SMART; SM00180; EGF_Lam; 3.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 3.
 KW Laminin EGF-like domain.
 FT NON TER 1
 FT 569 569
 SQ SEQUENCE 569 AA; 64479 MW; 67B1062E96E8B90 CRC64;
 Query Match 17.0%; Score 186; DB 13; Length 569;
 Best Local Similarity 27.4%; Pred. No. 1.4e-09;
 Matches 58; Conservative 38; Mismatches 92; Indels 24; Gaps 9;
 QY 5 RYCIPNDAGTADNRVRLN----PEAHLPSFVND-NDVG--TSWVSNVFTNITQLN 57
 DB 41 RYCVVTEKEDGRFRNCHCNMSDAKRAHPSPFLTLNPHNLTCWQSE---NYIQYQNV 97
 QY 58 TISVDLENGQYQVFIIOFFSPQPTETRIQRKKNLSLDWEDWQYFARNC-GAPGKNG 116
 DB 98 TLTLSL-GKKPEVTYVLSQFCSPRESMAIPKSMYDGKSWPFPYYSQCRMYNKPKA 156
 QY 117 DIEKPSVNLQLSNFT---PYSRGNVTFSLTPGNVPRPGYNNFNTPSLOESVKATQI 173
 DB 157 IITKQNEQAICTSDHTDHPHLSGGLIAFSTL----DGRPSAHDNDSPVLQDWVTATDI 212
 QY 174 ----RPFHFGQYTTTAVNLHRHYAYDEI 200
 DB 213 KVAPSLRHTFGDENEDDSLSARDSYFYAVS 244
 RESULT 7
 ID O42140 PRELIMINARY; PRT; 603 AA.
 AC O42140;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Netrin-1a.
 GN NTN1A.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97454757; PubMed=9268507;
 RA Lauderdale J.D., Davis N.M., Kuwada J.Y.;
 RT "Axon tracts correlate with netrin-1 expression in the zebrafish

embryo.";
 RL Mol. Cell. Neurosci. 9:293-313(1997).
 DR EMBL; AF002717; AAC60252.1; --
 DR HSSP; P02468; 1KLO.
 DR ZFIN; ZDB-GENE-990415-169; ntnla.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001886; LamNT.
 DR InterPro; IPR001134; Netrin_C.
 DR Pfam; PF00053; laminin_EGF; 3.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR Pfam; PF01759; NTR; 1.
 DR PRINTS; PR00011; EGFLAMININ.
 DR ProDom; PD002082; LamNT; 1.
 DR SMART; SM00643; C345C; 1.
 DR SMART; SM00180; EGF_Lam; 3.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 3.
 KW Laminin EGF-like domain.
 SQ SEQUENCE 603 AA; 67681 MW; 82EDB32E3E383F37 CRC64;
 Query Match 16.9%; Score 185; DB 13; Length 603;
 Best Local Similarity 27.0%; Pred. No. 1.9e-09;
 Matches 58; Conservative 41; Mismatches 86; Indels 30; Gaps 10;
 QY 5 RYCIPNDAGD-----TADNRVSKLNPEAHLPSFVND-NDVG--TSWVSNVFTNITQLN 54
 DB 75 RYCVVTEKEDGRHRNCHTCDASDPKKN---HPPAYLTDLNPHNLTCWQSD---NVLQYP 128
 QY 55 QGVTSVDLENGQYQVFIIOFFSPQPTETRIQRKKNLSLDWEDWQYFARNC-GAPGK 113
 DB 129 QNVTLTSL-GKKPEVTYVLSQFCSPRESMAIPKSMYDGKSWPFPYYSQCRMYNKP 187
 QY 114 NGDLEKPSVNLQLSNFT---PYSRGNVTFSLTPGNVPRPGYNNFNTPSLOESVK 170
 DB 198 SKAITKQNEQAICTSDHTDHPHLSGGLIAFSTL----DGRPSAHDNDSPVLQDW 243
 QY 171 TOI-----RPFHFGQYTTTAVNLHRHYAYDEI 200
 DB 244 TDIKVTFSLRHTFGDENEDDSLSARDSYFYAVS 278
 RESULT 8
 ID Q92429 PRELIMINARY; PRT; 604 AA.
 AC Q92429;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Netrin-1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21267048; PubMed=11356879;
 RA Manitt C., Colicos M.A., Thompson K.M., Rousselle E., Peterson A.C.,
 RA Kennedy T.E.;
 RT "Widespread Expression of Netrin-1 by Neurons and Oligodendrocytes in
 RT the Adult Mammalian Spinal Cord.";
 RL J. Neurosci. 21:3911-3922(2001).
 DR EMBL; AY028417; AAK17014.1; --
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001886; LamNT.
 DR InterPro; IPR001134; Netrin_C.
 DR Pfam; PF00053; laminin_EGF; 3.
 DR Pfam; PF01759; NTR; 1.
 DR PRINTS; PR00011; EGFLAMININ.
 DR ProDom; PD002082; LamNT; 1.

[illegible]

```
Db 190 SSKRDCLERFGQTLERITRDAAICTEYSRIVPLENGEIVVSLV----NRPGNMWPS 245
Qy 160 NTPSLOESVKAQIRF-----HFGQYVTTTAVNLRHRYVAVDEIT 201
Db 246 YSPLLREFTKATNVRFLRNTLLGHLGKALRDPVT-RRYYYSIKDIS 295

RESULT 11
Q9NL27 PRELIMINARY; PRT; 600 AA.
AC Q9NL27;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Netrin.
GN CI-NET1B.
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Clonidae; Clona.
OX NCBI_TaxID=7719;
RN [1]
RA Takamura K.;
RT "Expression patterns of ascidian netrin homologues.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB033332; BAA94303.1; -.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001886; LamNT.
DR Pfam; PF00053; laminin_EGF; 2.
DR Pfam; PF00055; laminin_Nterm; 1.
DR SMART; SM00136; LamNT; 1.
DR SMART; SM00180; EGF Lam; 2.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR KW Laminin EGF-like domain.
SQ SEQUENCE 600 AA; 68312 MW; 227D53324D17CCFB CRC64;

Query Match 15.0%; Score 163.5; DB 5; Length 600;
Best Local Similarity 27.1%; Pred. No. 2.4e-07;
Matches 48; Conservative 35; Mismatches 75; Indels 19; Gaps 7;

Qy 17 DNRVSL-----NPEAHLSP---VNDNDVGTSMVSNVFTNITQLNQGVTSVDLENG 66
Db 68 ENTSVRCVCEATGRSHPASLYTDINKHNLTYWQSKTFQSGEDQNRVELTISFEK- 126
Qy 67 QYQVYIIIFSPQPTETIRIQKKNLSLDWEDWQYFARNC-GAF--GMKNNGDLEKPS 123
Db 127 EYEISVIYQYSPRAAMIYKSMNHGRTWPYQYIAENCLRRFHKPKYKREANETNEQ 186
Qy 124 VNCLO-LSNFTPYSGNVTFSILTTPGPNYRPGYNNFTPSLOESVKATQIRPFHFG 179
Db 187 VLCSDFNSLYPFSLGALVFN----PKGRPSDDFEHSPILQNWVWTATDIKIVLAG 239
SQ SEQUENCE 600 AA; 68312 MW; 227D53324D17CCFB CRC64;

RESULT 12
Q9NL28 PRELIMINARY; PRT; 650 AA.
AC Q9NL28;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Netrin.
GN CI-NET1A.
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Clonidae; Clona.
OX NCBI_TaxID=7719;
```

```
RN [1]
RA Takamura K.;
RT "Expression patterns of ascidian netrin homologues.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB033331; BAA94302.1; -.
DR HSP; P02468; IKLO.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001886; LamNT.
DR InterPro; IPR001134; Netrin C.
DR Pfam; PF00053; laminin_EGF; 3.
DR Pfam; PF00055; laminin_Nterm; 1.
DR Pfam; PF01759; NTR; 1.
DR PRINTS; PR00011; EGFLAMININ.
DR ProDom; PD002082; LamNT; 1.
DR SMART; SM00643; C345C; 1.
DR SMART; SM00180; EGF Lam; 3.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR KW Laminin EGF-like domain.
SQ SEQUENCE 650 AA; 73807 MW; F5320C163AFB6E2C CRC64;

Query Match 15.0%; Score 163.5; DB 5; Length 650;
Best Local Similarity 27.1%; Pred. No. 2.6e-07;
Matches 48; Conservative 35; Mismatches 75; Indels 19; Gaps 7;

Qy 17 DNRVSL-----NPEAHLSP---VNDNDVGTSMVSNVFTNITQLNQGVTSVDLENG 66
Db 68 ENTSVRCVCEATGRSHPASLYTDINKHNLTYWQSKTFQSGEDQNRVELTISFEK- 126
Qy 67 QYQVYIIIFSPQPTETIRIQKKNLSLDWEDWQYFARNC-GAF--GMKNNGDLEKPS 123
Db 127 EYEISVIYQYSPRAAMIYKSMNHGRTWPYQYIAENCLRRFHKPKYKREANETNEQ 186
Qy 124 VNCLO-LSNFTPYSGNVTFSILTTPGPNYRPGYNNFTPSLOESVKATQIRPFHFG 179
Db 187 VLCSDFNSLYPFSLGALVFN----PKGRPSDDFEHSPILQNWVWTATDIKIVLAG 239
SQ SEQUENCE 650 AA; 73807 MW; F5320C163AFB6E2C CRC64;

RESULT 13
P91904 PRELIMINARY; PRT; 3704 AA.
AC P91904;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Laminin alpha (Epi-1 protein).
GN K08C7.3 OR EPI-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RA [1]
SQ SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Joh K., Hedgecock E.M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RA [2]
SQ SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Joh K., Zhu K., Hedgecock E.M.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RA [3]
SQ SEQUENCE FROM N.A.
RA Berks M.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB001074; BAA19229.1; -.
DR EMBL; AB016806; BAA32347.1; -.
DR EMBL; 270286; CAB61016.1; -.
DR HSP; Q92956; LUMA.
DR WormPep; K08C7.3; CE25049.
```



```

Query Match      14.7%; Score 160; DB 4; Length 529;
Best Local Similarity 26.2%; Pred. No. 4.4e-07;
Matches 58; Conservative 34; Mismatches 83; Indels 46; Gaps 11;

QY 3 AQRYPNDAGDTADNRVSRNLNPEA-HPLSFVND---NDVGTSMVSNVFTNITQLNQGYT 58
Db 70 AGAHCORCDAAD-----PORHNASYLTDFHSQDESTWQSPSWAFGVQYPTSVN 119
QY 59 ISVDLENGQYQVYIIIOFFSPOPTTEIRIQKKNESLWEDWQYFARNCGAFGMKRNGLD 118
Db 120 ITRLGKA-YEITYVRLKFTSRPESFALYKRSRADGPWEYQFYASAC---QKTYG-- 172
QY 119 EKPDSVNCI-----OLSNPTPYSGNVTPSILTPGNVRCYNNFYNTPSIQ 165
Db 173 -RPEG-QCLRPGEDERVAFTSEFSDISPLSGNVAFSTL-----EGRPSAYNFESPGIQ 226
QY 166 ESKATQI-----RFHFHGQYTTTAVNLRHRYTAVDEIT 201
Db 227 EWTSTELLISLDRLNTFGDDIFKPKV-LQSYYYAVSDFS 266

```

Search completed: December 31, 2003, 12:03:52
Job time : 12.9842 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2003, 11:57:49 ; Search time 1.2569 Seconds
(without alignments)
4722.097 Million cell updates/sec

Title: US-09-970-318-1
Perfect score: 126
Sequence: 1 QAPPQPGPTVWKISPTLRIE 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archesp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	126	100.0	1461	11 Q9JLP3	Q9JLP3 mus musculus
2	108	85.7	1512	11 Q8K3K1	Q8K3K1 rattus norv
3	71	56.3	1546	4 Q75445	Q75445 homo sapien
4	71	56.3	1546	4 Q9NS27	Q9NS27 homo sapien
5	54.5	43.3	4340	2 Q30764	Q30764 streptomyce
6	53	42.1	434	16 Q8XSJ0	Q8XSJ0 escherichia
7	52.5	41.7	1286	5 Q9TXR6	Q9TXR6 caenorhabdi
8	51	40.5	1243	12 Q8JFK6	Q8JFK6 heliothis z
9	50	39.7	647	3 Q8WZL8	Q8WZL8 varrovia li
10	50	39.7	946	12 Q59139	Q59139 human herpe
11	50	39.7	981	11 Q8CB00	Q8CB00 mus musculu
12	50	39.7	981	11 Q8CB81	Q8CB81 mus musculu
13	50	39.7	989	11 Q99ML9	Q99ML9 mus musculu
14	50	39.7	1133	3 Q8NIT4	Q8NIT4 neurospora
15	49.5	39.3	564	11 Q8K4M2	Q8K4M2 rattus norv
16	49.5	39.3	570	11 Q99J80	Q99J80 rattus norv

17	49	38.9	286	16 Q8YXG6	Q8YXG6 anabaena sp
18	49	38.9	318	4 Q8NFG7	Q8NFG7 homo sapien
19	49	38.9	321	11 Q9QXFS	Q9QXFS mesocricetu
20	49	38.9	328	4 Q8N6A6	Q8N6A6 homo sapien
21	49	38.9	328	4 Q8N6A7	Q8N6A7 homo sapien
22	49	38.9	332	6 Q9N257	Q9N257 oryctolagus
23	49	38.9	335	4 Q96106	Q96106 homo sapien
24	49	38.9	338	4 Q8NFD3	Q8NFD3 homo sapien
25	49	38.9	356	4 Q9HAA9	Q9HAA9 homo sapien
26	49	38.9	364	10 Q8S0C5	Q8S0C5 oryza sativ
27	49	38.9	394	17 Q9HLM3	Q9HLM3 thermoplas
28	49	38.9	438	16 Q9K4A7	Q9K4A7 streptomyce
29	49	38.9	1135	13 Q9YHD3	Q9YHD3 xenopus lae
30	49	38.9	3164	12 Q6S088	Q6S088 human herpe
31	48.5	38.5	521	4 Q8TEN0	Q8TEN0 homo sapien
32	48.5	38.5	778	4 Q8WY92	Q8WY92 homo sapien
33	48.5	38.5	834	4 Q81Z84	Q81Z84 homo sapien
34	48	38.1	216	11 Q9CWE8	Q9CWE8 mus musculu
35	48	38.1	290	12 Q9Q8U2	Q9Q8U2 shope fibro
36	48	38.1	333	11 Q91Y74	Q91Y74 mus musculu
37	48	38.1	333	11 Q92IR5	Q92IR5 mus musculu
38	48	38.1	333	11 P97354	P97354 m cmp-n-ace
39	48	38.1	370	10 Q64895	Q64895 arabidopsis
40	48	38.1	396	10 Q9SJF3	Q9SJF3 arabidopsis
41	48	38.1	421	5 Q9VM68	Q9VM68 drosophila
42	48	38.1	487	16 Q8XV67	Q8XV67 raistonia s
43	48	38.1	652	3 Q9Y835	Q9Y835 penicillium
44	48	38.1	838	5 Q9VQA9	Q9VQA9 drosophila
45	48	38.1	2338	5 Q94269	Q94269 caenorhabdi

ALIGNMENTS

RESULT 1
Q9JLP3 PRELIMINARY; PRT; 1461 AA.

ID Q9JLP3
AC Q9JLP3
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Putative extracellular matrix protein MUSH2A.
GN USH2A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rudy J.D., Talmadge C.B., Weston M.D., Yao S.-F., Cosgrove D.,
RA Ahmad I., Kimberling W., Sumegi J.;
RT "Isolation and Characterization of the Murine Homolog of the Usher
RT Syndrome Type 2A Gene."
RL Submitted (MAY-1999) to the ENBL/GenBank/DBJ databases.
DR EMBL; AF151717; RAP70550.1;
DR HSSP; P02468; ITLE.
DR MGD; MGI:1341292; Ush2a.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001886; Laminin_EGF.
DR Pfam; PF00041; fn3; 4.
DR Pfam; PF00053; laminin_EGF; 10.
DR Pfam; PF00055; laminin_Nterm; 1.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SM00180; EGF_Lam; 8.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00136; Laminin; 1.
DR PROSITE; PS00022; EGF 1; 7.
DR PROSITE; PS01248; LAMININ TYPE EGF; 7.
KW Laminin EGF-like domain; Matrix protein.
SQ SEQUENCE 1461 AA; 160989 MW; 114436D877C38A15 CRC64;

```

Query Match      100.0%; Score 126; DB 11; Length 1461;
Best Local Similarity 100.0%; Pred. No. 4.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QAPPQTQGPPTVWKISPTLRIE 23
    |||||
DB 1237 QAPPQTQGPPTVWKISPTLRIE 1259
    |||||

RESULT 2
Q8K3K1 ID Q8K3K1 PRELIMINARY; PRT; 1512 AA.
AC Q8K3K1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Usherin.
GN USH2A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22150873; PubMed=12160733;
RA Huang D., Eudy J.D., Uzvolgyi E., Davis J.R., Talmadge C.B.,
RA Pretto D., Weston M.D., Lehman J.E., Zhou M., Seemayer T.A., Ahmad I.,
RA Kimberling W.J., Sumegi J.;
RT "Identification of the Mouse and Rat Orthologs of the Gene Mutated in
RT Usher Syndrome Type IIA and the Cellular Source of USH2A mRNA in
RT Retina, a Target Tissue of the Disease.";
RL Genomics 80:195-203(2002).
DR EMBL: AY077844; AAL78289.1; -.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR006558; LamG like.
DR InterPro; IPR002049; Laminin_EGF.
DR Pfam; PF00041; fn3; 4.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00560; FN3; 4.
DR SMART; SM00060; LamGL; 1.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 6.
KW Laminin EGF-like domain.
SQ SEQUENCE 1512 AA; 167591 MW; E37D79F5D00044C1 CRC64;

Query Match      85.7%; Score 108; DB 11; Length 1512;
Best Local Similarity 82.8%; Pred. No. 2.1e-07;
Matches 19; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 QAPPQTQGPPTVWKISPTLRIE 23
    |||||
DB 1236 QAPPQTQGPPTVWKISPTLRIE 1258
    |||||

RESULT 3
Q75445 ID Q75445 PRELIMINARY; PRT; 1546 AA.
AC Q75445;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Usher syndrome type IIA protein.
GN USH2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=98288356; PubMed=9624053;
RA Eudy J.D., Weston M.D., Yao S.F., Hoover D.M., Rehm H.L., Ahmad I.,
RA Ma-Edmonds M., Yan D., Cheng J.J., Beisel K.W., Ayuso C., Cremers C.,
RA Davenport S., Moller C., Talmadge C.B., Tamayo M., Swaroop A.,
RA Morton C.C., Kimberling W.J., Sumegi J.;
RT "Mutation of a gene encoding a protein with extracellular matrix
RT motifs in Usher syndrome type IIA.";
RL Science 280:1753-1757(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Eudy J.D., Yao S.F., Cheng J.J., Weston M.D., Sumegi J.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF055580; AAC23748.2; -.
DR HSSP; P02468; IKLO.
DR Genew; HGNC:12601; USH2A.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR006558; LamG like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001886; LamNT.
DR Pfam; PF00041; fn3; 4.
DR Pfam; PF00053; laminin_Nterm; 1.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00180; EGF_Lam; 10.
DR SMART; SM00560; FN3; 4.
DR SMART; SM00060; LamGL; 1.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 7.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
KW Laminin EGF-like domain.
SQ SEQUENCE 1546 AA; 170985 MW; 49CD0A95A614959F CRC64;

Query Match      56.3%; Score 71; DB 4; Length 1546;
Best Local Similarity 60.9%; Pred. No. 0.072;
Matches 14; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 QAPPQTQGPPTVWKISPTLRIE 23
    |||||
DB 1240 QAPPQTQGPPTVWKISPTLRIE 1262
    |||||

RESULT 4
Q9NS27 ID Q9NS27 PRELIMINARY; PRT; 1546 AA.
AC Q9NS27;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Usher syndrome type IIA protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20206315; PubMed=10729113;
RA Weston M.D., Eudy J.D., Fugita S., Yao S.-F., Usami S., Cremers C.,
RA Greenburg J., Ramesar R., Martini A., Moller C., Smith R.J.,
RA Sumegi J., Kimberling W.J.;
RT "Genomic structure and identification of novel mutations in usherin,
RT the gene responsible for Usher syndrome type IIA.";
RL Am. J. Hum. Genet. 66:1199-1210(2000).
DR EMBL; AF091889; AAF75819.1; -.
DR EMBL; AF091873; AAF75819.1; JOINED.
DR EMBL; AF091874; AAF75819.1; JOINED.
DR EMBL; AF091875; AAF75819.1; JOINED.
DR EMBL; AF091876; AAF75819.1; JOINED.
DR EMBL; AF091877; AAF75819.1; JOINED.
DR EMBL; AF091878; AAF75819.1; JOINED.
DR EMBL; AF091879; AAF75819.1; JOINED.
DR EMBL; AF091880; AAF75819.1; JOINED.
DR EMBL; AF091881; AAF75819.1; JOINED.

```

DR EMBL; AF091882; AAF75819.1; JOINED.
 DR EMBL; AF091883; AAF75819.1; JOINED.
 DR EMBL; AF091884; AAF75819.1; JOINED.
 DR EMBL; AF091885; AAF75819.1; JOINED.
 DR EMBL; AF091886; AAF75819.1; JOINED.
 DR EMBL; AF091887; AAF75819.1; JOINED.
 DR EMBL; AF091888; AAF75819.1; JOINED.
 DR HSSP; P02468; 1KLO.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR006558; LamG like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001886; LamNT.
 DR Pfam; PF00041; fn3; 4.
 DR Pfam; PF00053; laminin_EGF; 10.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR SMART; SM00180; EGF_Lam; 10.
 DR SMART; SM00060; FN3; 4.
 DR SMART; SM00560; LamGL; 1.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 7.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
 KW Laminin EGF-like domain.
 SQ SEQUENCE 1546 AA; 171001 MW; 64BD435996C1E17B CRC64;

Query Match 56.3%; Score 71; DB 4; Length 1546;
 Best Local Similarity 60.9%; Pred. No. 0.072;
 Matches 14; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 QAPPQTQGPPTVWKISPTLEIRIE 23
 ||||| : : : : :
 DB 1240 QAPPRLSPPMQKISSTELHVE 1262

RESULT 5
 O30764

ID O30764 PRELIMINARY; PRT; 4340 AA.
 AC O30764;
 DT 01-JAN-1998 (T-EMBLrel. 05, Created)
 DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Polypeptide synthase modules 1 and 2.
 GN NIDAI.
 OS Streptomyces caelestis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=36816;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRL-2821;
 RX MEDLINE=98053867; PubMed=9393718;
 RA Kakavas S.J.; Katz L.; Stassi D.;
 RT Identification and characterization of the midamycin polyketide
 synthase genes from Streptomyces caelestis.;
 RL J. Bacteriol. 179:7515-7522(1997).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 (SDR) FAMILY.
 CC EMBL; AF016585; AAC46024.1; -;
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR002198; ADH short.
 DR InterPro; IPR004410; FabD
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR006162; Ppantne attach.
 DR InterPro; IPR006163; Pp bind.
 DR Pfam; PF00698; Acyl trans; 3.
 DR Pfam; PF00106; adh short; 2.
 DR Pfam; PF00109; ketoacyl-synt; 3.
 DR Pfam; PF00550; pp-binding; 3.
 DR TIGRPFAMs; TIGR00128; fabD; 3.
 DR PROSITE; PS50075; ACP DOMAIN; 3.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.

DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
 KW Oxidoreductase; Phosphopantetheine; Transferase.
 SQ SEQUENCE 4340 AA; 457589 MW; D59A734CB5FB795D CRC64;

Query Match 43.3%; Score 54.5; DB 2; Length 4340;
 Best Local Similarity 46.4%; Pred. No. 63;
 Matches 13; Conservative 4; Mismatches 6; Indels 5; Gaps 2;

QY 1 QAPPQTQGPPTV--WKISPTLEIRIE 23
 : : : : :
 DB 1438 EAPARTQAPPVIAPLSAHTPTALRAQ 1465

RESULT 6

O8X5J0
 ID O8X5J0 PRELIMINARY; PRT; 434 AA.
 AC O8X5J0;
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Putative permease, hexosephosphate transport (Putative regulatory
 protein).
 DE PROSITE; PS0461 OR EGS0416.
 GN Z0461 OR EGS0416.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T.; Plunkett G. III; Burland V.; Mau B.; Glasner J.D.;
 RA Rose D.J.; Mayhew G.F.; Evans P.S.; Gregor J.; Kirkpatrick H.A.;
 RA Posfai G.; Hackett J.; Klink S.; Boutin A.; Shao Y.; Miller L.;
 RA Grobeck E.J.; Davis N.W.; Lim A.; Dimalanta E.T.; Potamouisis K.;
 RA Apodaca J.; Anantharaman T.S.; Lin J.; Yen G.; Schwartz D.C.;
 RA Welch R.A.; Blattner P.R.;
 RT Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.*;
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T.; Makino K.; Ohnishi M.; Kurokawa K.; Ishii K.; Yokoyama K.;
 RA Han C.-G.; Ohtsubo E.; Nakayama K.; Murata T.; Tanaka M.; Tobe T.;
 RA Iida T.; Takami H.; Honda T.; Sasakawa C.; Ogasawara N.; Yasunaga T.;
 RA Kuhara S.; Shiba T.; Hattori M.; Shinagawa H.;
 RT Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.*;
 RL DNA Res. 8:11-22(2001).
 DR EMBL; AF005215; AAG54712.1; -;
 DR EMBL; AP002551; BAB33839.1; -;
 DR InterPro; IPR007114; MFS.
 DR PROSITE; PS50850; MFS; 1.
 KW Complete proteome.
 SQ SEQUENCE 434 AA; 47920 MW; 67B819BA39EFD0E8 CRC64;

Query Match 42.1%; Score 53; DB 16; Length 434;
 Best Local Similarity 54.5%; Pred. No. 9.3;
 Matches 12; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

QY 4 PQQTQGPPTV--WKISPTLEIRIE 23
 : : : : :
 DB 198 PQEGLPTVGHWRHDPLELRQE 219

RESULT 7

O9TXR6
 ID O9TXR6 PRELIMINARY; PRT; 1286 AA.
 AC O9TXR6;
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 DT 01-MAY-2003 (T-EMBLrel. 23, Last annotation update)

DE	M0E10.2 protein.
GN	M0E10.2.
OS	Caeenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
CC	Rhabditiidae; Peloderinae; Caeonorhabditis.
OX	NCB_I_TaxID=6239;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Bristol N2;
RX	MEDLINE=99069613; PubMed=9851916;
RA	None;
RT	"genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.";
RL	Science 282:2012-2018(1998).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Bristol N2.
RA	Geisel C., Kramer J., Smith A.;
RT	"The sequence of C. elegans cosmid M0E10.1.";
RL	Submitted [OCT-1998] to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Bristol N2;
CA	Waterston R.;
RL	Submitted [JUN-1999] to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF099921; AAC68807.1; -.
DR	WormPep; M0E10.2; CE19534.
DR	InterPro; IPR001507; Endoglin/CDI05.
DR	InterPro; IPR002865; P-rich_extensn.
DR	InterPro; IPR002035; VWF_A.
DR	Pfam; PF00092; vwa; 1.
DR	Pfam; PF00100; zona_pellucida; 1.
DR	PRINTS; PR01217; PRICHEXTNSN.
DR	PRINTS; PRO0453; VWFADOMAIN.
DR	SMART; SMART0327; VWA; 1.
DR	SMART; SMART0241; ZP; 1.
DR	DROSITE; PS50234; VWFA; 1.
SQ	SEQUENCE 1286 AA; 140794 MW; 95781989BC127CCE CRC64;
Query Match	41.7%; Score 52.5; DB 5; Length 1286;
Best Local Similarity	50.0%; Pred.No.35;
Matches	11; Conservative 3; Mismatches 7; Indels 1; Gaps
Qy'	1 QAPPQTGGPTTWKISPTELRI 22 :: : :: :
Db	336 EAAPTQTTPTSW-TRPTDGRV 356 :: : :: :
RESULT 8	
ID	'QBKK6 PRELMINARY; PRT: 1243 AA.
AC	BQJKK6;
DT	01-OCT-2002 (TReMBlRel. 22, Created)
DT	01-OCT-2002 (TReMBlRel. 22, Last sequence update)
DT	01-MAR-2003 (TReMBlRel. 23, Last annotation update)
DN	Mucin-like glycoprotein 900.
GN	ORF107.
OS	Heliobacterium salinarum virus 1.
OC	Viruses; unclassified viruses.
OX	NCBI_TaxID=29250;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=22174892; PubMed=12186886;
RA	Cheng C.H.; Liu S.M.; Chow T.Y.; Hsiao Y.X.; Wang D.P.; Huang J.J.;
CH	Chen H.H.;
RT	"Analysis of the Complete Genome Sequence of the Hz-1 Virus Suggests
RL	that It Is Related to Members of the Baculoviridae.";
J. Virol.	76:9024-9034(2002).
RN	[2]
RP	SEQUENCE FROM N.A.
RA	Cheng C.H.; Liu H.M.; Hsiao Y.Y.; Chow T.Y.; Chen H.H.;
RL	Submitted [JUN-2001] to the EMBL/GenBank/DBJ databases.
RN	[3]

```

RP SEQUENCE FROM N.A.
RA Cheng C.H., Hsiao Y.Y., Liu S.M., Chow T.Y., Chen H.H.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF451898; AAN04401.1; -.
DR InterPro: IPR002965; P-rich extensin.
DR PRINTS: PR01217; PRICHEXTENSIN.
SQ SEQUENCE 1243 AA; 136853 MW; B5400BD894B5CBA9 CRC64;

Query Match 40.5%; Score 51; DB 12; Length 1243;
Best Local Similarity 47.8%; Pred. No. 56;
Matches 11; Conservative 4; Mismatches 4; Indels 4; Gaps 1;

QY 3 PPQTQGPPTV---WKISPTEL 21
DB 830 PPLTQPPTTINLTTFKLSPTTLK 852
|||||:|||||:
| | | | | : | | | | | :

RESULT 9
Q8WZL8
ID Q8WZL8 PRELIMINARY; PRT; 647 AA.
AC Q8WZL8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Transcriptional repressor, TUP1.
GN YUP1.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=POLA;
RA Cabello J., Dominguez A.;
RT "The Yarrowia lipolytica TUP1, a gene involved in the yeast-to-hypha
RT transition."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=POLA;
RA Cabello J.;
RT "New strategies for studying dimorphism. Isolation and characterization
RT of a transcriptional repressor, YITUP1, from Yarrowia lipolytica.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ252128; CAC81004.1; -.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 7.
DR PRINTS: PR00320; GPROTEINREP.T.
DR ProDom: PD000018; WD40; 6.
DR SMART: SM00320; WD40; 7.
DR PROSITE: PS00678; WD_REPEATS_1; 4.
DR PROSITE: PS50082; WD_REPEATS_2; 6.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 647 AA; 71292 MW; 7A7CD9C988A35DC5 CRC64;

Query Match 39.7%; Score 50; DB 3; Length 647;
Best Local Similarity 50.0%; Pred. No. 40;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 QAPPQTQGPPTVWKISPTEL 20
DB 279 QGPPQQQPPQQQQSPREM 298
|||||:|||||:
| | | | | : | | | | | :

RESULT 10
Q69139
ID Q69139 PRELIMINARY; PRT; 946 AA.
AC Q69139;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nuclear antigen EBNA-3B.

```

OS Human herpesvirus 4 (Epstein-Barr virus).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=10376;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90347798; PubMed=2166806;
 RA Sample J., Young L., Martin B., Chatman T., Kieff B.D, Rickinson A.;
 RT "Epstein-Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B and
 RT EBNA-3C genes.";
 RL J. Virol. 64:4084-4092(1990).
 DR EMBL; M34440; AAA45894.1; -;
 SQ SEQUENCE 946 AA; 104103 MW; 7F428D409134B869 CRC64;

Query Match 39.7%; Score 50; DB 12; Length 946;
 Best Local Similarity 52.9%; Pred. No. 60;
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 PQQTGGPPTVWKISPTL 20
 Db 753 PPQGGPPTVWKISPTL 769

RESULT 11
 Q8C8C0 ID Q8C8C0 PRELIMINARY; PRT; 981 AA.
 AC Q8C8C0;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Arkadia.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 DR EMBL; AK036351; BAC29394.1; -;
 SQ SEQUENCE 981 AA; 106917 MW; 14368A4582C1F5D5 CRC64;

Query Match 39.7%; Score 50; DB 11; Length 981;
 Best Local Similarity 62.5%; Pred. No. 62;
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 APPQTGGPPTVWKISP 17
 Db 666 APPQTGGPPTVWKISP 681

RESULT 12
 Q8C881 ID Q8C881 PRELIMINARY; PRT; 981 AA.
 AC Q8C881;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Arkadia.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.

RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 DR EMBL; AK048110; BAC33245.1; -;
 SQ SEQUENCE 981 AA; 106928 MW; E421C32F1A64CFEE CRC64;

Query Match 39.7%; Score 50; DB 11; Length 981;
 Best Local Similarity 62.5%; Pred. No. 62;
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 APPQTGGPPTVWKISP 17
 Db 666 APPQTGGPPTVWKISP 681

RESULT 13
 Q99ML9 ID Q99ML9 PRELIMINARY; PRT; 989 AA.
 AC Q99ML9;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Arkadia.
 GN ARK OR ARKADIA.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/Sv;
 RX MEDLINE=21195971; PubMed=11298452;
 RA Episkopou V., Arkell R., Timmons P.M., Walsh J.J., Andrew R.L.,
 RA Swan D.;
 RT "Induction of the mammalian node requires Arkadia function in the
 RT extraembryonic lineages";
 RL Nature 410:825-830(2001).
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 DR EMBL; AF330197; AAK38272.1; -;
 DR MGD; MGI:1934919; Ark.
 DR InterPro; IPR001211; PhospholipaseA2.
 DR Pfam; PF00097; zf-C3HC4_1.
 DR SMART; SM00184; RING; 1.
 DR PROSITE; PS00118; PA2_HIS; 1.
 DR PROSITE; PS50089; ZF_RING_2; 1.
 KW Metal-binding; Zinc; Zinc-finger.
 SQ SEQUENCE 989 AA; 107896 MW; 212E3C37BC70DCB5 CRC64;

Query Match 39.7%; Score 50; DB 11; Length 989;
 Best Local Similarity 62.5%; Pred. No. 62;
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 APPQTGGPPTVWKISP 17
 Db 666 APPQTGGPPTVWKISP 681

RESULT 14
 Q8NIT4 ID Q8NIT4 PRELIMINARY; PRT; 1133 AA.
 AC Q8NIT4;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN B13H18.270.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]

```

RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL807374; CAD37073.1; -
KW Hypothetical protein.
SQ SEQUENCE 1133 AA; 124593 MW; 7B0A9C52CF59DF17 CRC64;

Query Match          39.7%; Score 50; DB 3; Length 1133;
Best Local Similarity 50.0%; Pred. No. 72;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 PPOTQGPPTVWKISPT 18
   ||| ||| ||| |||
DB 446 PTPQSPPTLYRHSPS 461

RESULT 15
Q8K4M2 PRELIMINARY; PRT; 564 AA.
AC Q8K4M2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE FAD-dependent sulphydryl oxidase.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Benayoun B., Ennard-Reve A., Ennard P.;
RT "Rat FAD-dependent sulphydryl oxidase (SOX) mRNA."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF217799; AAM67412.1; -
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR006662; ThioRed.
DR InterPro; IPR006683; ThioRedox_dom2.
DR Pfam; PF00085; ThioRed; 1.
DR PRINTS; PR00421; THIOREDOXIN.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
KW Redox-active center.
SQ SEQUENCE 564 AA; 62856 MW; 33B0A60AE6AF1A84 CRC64;

Query Match          39.3%; Score 49.5; DB 11; Length 564;
Best Local Similarity 43.5%; Pred. No. 41;
Matches 10; Conservative 2; Mismatches 4; Indels 7; Gaps 1;

QY 1 QAPPOTQGP-----PTWKIS 16
   :||| |
DB 269 EAPPTAAPVTPDKIAPTWKFA 291

```

Search completed: December 31, 2003, 12:03:50
 Job time : 2.2569 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 06:39:50 / Search time 14799 Seconds
(without alignments)
17498.340 Million cell updates/sec

Title: US-09-970-318-3
Perfect score: 6330
Sequence: 1 tggctgctgcagaatact.....gcaaaaaaaaaaaaaaaa 6330

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em.sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	6307.6	99.6	6332	9	AF055580	AF055580 Homo sapi
2	2748	43.4	5126	10	AF151717	AF151717 Mus muscu
3	2683.6	42.4	6146	10	AY077844	AY077844 Rattus no
4	1546.2	24.4	2652	9	HSUSH2A17	AF091889 Homo sapi
5	1546.2	24.4	115940	9	AC119429	AC119429 Homo sapi
6	1546.2	24.4	163815	9	AC138024	AC138024 Homo sapi
7	1531	24.2	184849	2	AC021209	AC021209 Homo sapi
8	683.4	10.8	2749	9	HSUSH2A01	AF091873 Homo sapi
9	645.8	10.2	153705	9	AL445650	AL445650 Human DNA
10	645.8	10.2	1593	9	HSUSH2A10	AF091882 Homo sapi
11	645.8	10.2	184849	2	AC021209	AC021209 Homo sapi
12	645.8	10.2	290708	2	AL161841	AL161841 Homo sapi
13	498.4	7.9	2840	9	HSUSH2A14	AF091886 Homo sapi
14	476.4	7.5	2749	6	AX347325	AX347325 Sequence
15	437	6.9	2749	6	AX347324	AX347324 Sequence
16	432.4	6.8	152888	10	AC121892	AC121892 Mus muscu
17	427.6	6.8	241363	2	AC106298	AC106298 Rattus no
18	412.8	6.5	290708	2	AL161841	AL161841 Homo sapi
19	301	4.8	3175	9	HSUSH2A05	AF091877 Homo sapi
20	301	4.8	93856	9	AL358858	AL358858 Human DNA
21	257.6	4.1	257291	2	AC122122	AC122122 Mus muscu
22	250	3.9	317136	2	AC098150	AC098150 Rattus no
23	233.8	3.7	290002	2	AC098145	AC098145 Rattus no
24	204	3.2	317136	2	AC098150	AC098150 Rattus no
25	199.2	3.1	1348	9	HSUSH2A07	AF091879 Homo sapi
26	199	3.1	883	9	HSUSH2A09	AF091881 Homo sapi
27	185.2	2.9	1047	9	HSUSH2A11	AF091883 Homo sapi
28	178.6	2.8	106863	5	AL928842	AL928842 Zebrafish
29	176	2.8	1255	9	HSUSH2A02	AF091874 Homo sapi
30	174.8	2.8	1329	9	HSUSH2A15	AF091887 Homo sapi
31	166.4	2.6	1268	9	HSUSH2A12	AF091884 Homo sapi
32	162.4	2.6	1364	9	HSUSH2A13	AF091885 Homo sapi
33	162.2	2.6	269897	2	AC121799	AC121799 Mus muscu
34	155.8	2.5	726	11	BV028598	BV028598 S212P6051
35	147	2.3	1662	9	HSUSH2A16	AF091888 Homo sapi
36	136.8	2.2	948	9	HSUSH2A03	AF091875 Homo sapi
37	134.8	2.1	1243	9	HSUSH2A08	AF091880 Homo sapi
38	117.4	1.9	5433	6	AX045107	AX045107 Sequence
39	117.4	1.9	5433	6	AX045324	AX045324 Sequence
40	117.4	1.9	5433	6	AX045105	AX045105 Sequence
41	117.4	1.9	5613	6	AX045322	AX045322 Sequence
42	117.4	1.9	5613	6	AX395746	AX395746 Sequence
43	117.4	1.9	5613	6	AX409725	AX409725 Sequence
44	117.4	1.9	5613	6	AX463742	AX463742 Sequence
45	117.4	1.9	5613	6	AX463742	AX463742 Sequence

ALIGNMENTS

RESULT 1
AF055580
LOCUS AF055580 6332 bp mRNA linear PRI 21-MAR-2002
DEFINITION Homo sapiens Usher syndrome type IIA protein (USH2A) mRNA, complete cds.
ACCESSION AF055580
VERSION AF055580.2 GI:19584001
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 6332)
AUTHORS Eudy,J.D., Weston,M.D., Yao,S.F., Hoover,D.M., Rehm,H.L., Ahmad,I., Ma-Edmonds,M., Yan,D., Cheng,J.J., Beisel,K.W., Ayuso,C.,

Db 3181 AATCAGTGTCAACGAGGTTTTATATTTCTCAGGCAATGCCACTGGCTGCTGCCATGC 3240
Qy 3239 TCATGCCATACAACTGGCGCAGTTAATCACATCTGTAAATAGCCCTGACTGGTCAAGTGTGT 3298
Db 3241 TCATGCCATACAACTGGCGCAGTTAATCACATCTGTAAATAGCCCTGACTGGTCAAGTGTGT 3300
Qy 3299 TGCCAAAGATCTTCATCTGCTGGCGCAGCTGTGACCAATGCAAGACCAATTAATCTTTGGA 3358
Db 3301 TGCCAAAGATCTTCATCTGCTGGCGCAGCTGTGACCAATGCAAGACCAATTAATCTTTGGA 3360
Qy 3359 TTTGATCTCAGACTGGAAGATGTGAGCTTTGTAAATGTCACTCTCAGGAGCCCTTCAAT 3418
Db 3361 TTTGATCTCAGACTGGAAGATGTGAGCTTTGTAAATGTCACTCTCAGGAGCCCTTCAAT 3420
Qy 3419 GAACCTGTCTCAGCTGGTGTGACAGGCAAGTGTCTGTAATGCTCACTCTCAGGAGCCCTTCAAT 3478
Db 3421 GAACCTGTCTCAGCTGGTGTGACAGGCAAGTGTCTGTAATGCTCACTCTCAGGAGCCCTTCAAT 3480
Qy 3479 TGTGATCTGTGTTCCAGTGTGCAAGGCACTTGGATGTCAACATCTTAATTTGGGTGCGC 3538
Db 3481 TGTGATCTGTGTTCCAGTGTGCAAGGCACTTGGATGTCAACATCTTAATTTGGGTGCGC 3540
Qy 3539 AAAAATCCATTTCCAGCAACCTTCGCCCCCAGAGGCAAGTGTCAAAATGCTCTGCTATCAAT 3598
Db 3541 AAAAATCCATTTCCAGCAACCTTCGCCCCCAGAGGCAAGTGTCAAAATGCTCTGCTATCAAT 3600
Qy 3599 CTCTCTGGAGTCCACTGTGATCTTCCAAATGCCACTGGCTTACTTACAGTTTACTCAGG 3658
Db 3601 CTCTCTGGAGTCCACTGTGATCTTCCAAATGCCACTGGCTTACTTACAGTTTACTCAGG 3660
Qy 3659 GATGGTTTTGAAATCTACACAGAGGATCAATACCCATACAGTATTAATTAATCTTCTTA 3718
Db 3661 GATGGTTTTGAAATCTACACAGAGGATCAATACCCATACAGTATTAATTAATCTTCTTA 3720
Qy 3719 GACACAGACTGTGATCAATATACCAATATTTCTTATTAATTAATTAATTAATTAATTAAT 3778
Db 3721 GACACAGACTGTGATCAATATACCAATATTTCTTATTAATTAATTAATTAATTAATTAAT 3780
Qy 3779 GGTTCACAAAGAGTGTAGCTGTCACTTACAAAGCAAAACAGGGGTCCAGAGGGAAAC 3838
Db 3781 GGTTCACAAAGAGTGTAGCTGTCACTTACAAAGCAAAACAGGGGTCCAGAGGGAAAC 3840
Qy 3839 TTGACTTTAAGTTATATCATTTCTTATTTGGCTCAGACTCTGTGACATTAATCTGCAACA 3898
Db 3841 TTGACTTTAAGTTATATCATTTCTTATTTGGCTCAGACTCTGTGACATTAATCTGCAACA 3900
Qy 3899 CTCTCAAAATCAATCTGGTCCCATAGAGAAATATATTTTGTCTGTGCCCCCTTTGGCTGGT 3958
Db 3901 CTCTCAAAATCAATCTGGTCCCATAGAGAAATATATTTTGTCTGTGCCCCCTTTGGCTGGT 3960
Qy 3959 GGTTCAGCATGTGTTCTTCTTCAAGAGTGTATGAAACCTCAGTACCATCTGGAATCTGGTT 4018
Db 3961 GGTTCAGCATGTGTTCTTCTTCAAGAGTGTATGAAACCTCAGTACCATCTGGAATCTGGTT 4020
Qy 4019 CCATTTGCCAAGTACGATTTTCTTCTTACAGGCTGTGATAGCGGGGCTGTTTACACAGC 4078
Db 4021 CCATTTGCCAAGTACGATTTTCTTCTTACAGGCTGTGATAGCGGGGCTGTTTACACAGC 4080
Qy 4079 TTGCCCCATTAAGTACAGCAGCCAGGCCCCCTTCCCAAGAGCTTAAGTCCACCTAGATG 4138
Db 4081 TTGCCCCATTAAGTACAGCAGCCAGGCCCCCTTCCCAAGAGCTTAAGTCCACCTAGATG 4140
Qy 4139 CAGAAATCAGTTCTACAGAACTTCAATGTAGAAATGGTCTCCACAGCGGAACTAAATGGA 4198
Db 4141 CAGAAATCAGTTCTACAGAACTTCAATGTAGAAATGGTCTCCACAGCGGAACTAAATGGA 4200
Qy 4199 ATAAATTAAGATATGAATATATCATGAGAGAACTGAGATCTAATAAGAAACCAATCT 4258
Db 4201 ATAAATTAAGATATGAATATATCATGAGAGAACTGAGATCTAATAAGAAACCAATCT 4260
Qy 4259 GAGAAATGCGAGTTTTCAGACAGGTTGGTGGCTCAGTCTCTCAATTTGTAGAAATCG 4318
Db 4261 GAGAAATGCGAGTTTTCAGACAGGTTGGTGGCTCAGTCTCTCAATTTGTAGAAATCG 4320

Qy 4319 GCCAATGAAATGCAATTAATAAACCTCTCTCAAAACAATGACAAACCATCACTGGCTTGGAGCCA 4378
Db 4321 GCCAATGAAATGCAATTAATAAACCTCTCTCAAAACAATGACAAACCATCACTGGCTTGGAGCCA 4380
Qy 4379 TACACCAAGTATGAGTGTAGAGTCTTACAGTCTTGAATATATGGCTGGAAAGTGTCTTCTGCC 4438
Db 4381 TACACCAAGTATGAGTGTAGAGTCTTACAGTCTTGAATATATGGCTGGAAAGTGTCTTCTGCC 4440
Qy 4439 TGGGTCTCTCAAAAGAACGGGAGAAATCAGACCTGTATTAATCATATGATCCCTCTCAGTCTTT 4498
Db 4441 TGGGTCTCTCAAAAGAACGGGAGAAATCAGACCTGTATTAATCATATGATCCCTCTCAGTCTTT 4500
Qy 4499 CCCCTCTCTTCTGTPACTCTCTCAATATCTCTGGGAGAGCCAGCAGATTAATTTACAAGA 4558
Db 4501 CCCCTCTCTTCTGTPACTCTCTCAATATCTCTGGGAGAGCCAGCAGATTAATTTACAAGA 4560
Qy 4559 GGAAGATTTGGGGGTATGACATCAATATGCTTTCTGAAACAATCACTCAACAGTCTAT 4618
Db 4561 GGAAGATTTGGGGGTATGACATCAATATGCTTTCTGAAACAATCACTCAACAGTCTAT 4620
Qy 4619 CCCATGCGGTTTTTACAGCTGTGTCACACTGTCTAAATCCCAAGAACTATCTTACACTGTA 4678
Db 4621 CCCATGCGGTTTTTACAGCTGTGTCACACTGTCTAAATCCCAAGAACTATCTTACACTGTA 4680
Qy 4679 GAAGACTGGAACCTTTATAGGATATATAGTATTACTTATTAATCTCTGCAATTCAGTTGGT 4738
Db 4681 GAAGACTGGAACCTTTATAGGATATATAGTATTACTTATTAATCTCTGCAATTCAGTTGGT 4740
Qy 4739 TGTGTGACAGTGTCTTGGGAGCAGCAAACTTTTAGCAGCAGCAGCAGCAACCTGAGG 4798
Db 4741 TGTGTGACAGTGTCTTGGGAGCAGCAAACTTTTAGCAGCAGCAGCAGCAACCTGAGG 4800
Qy 4799 CCACCTCTGGTTAAAGGAATCAACAGCACAACCAATCCATCTTAAGTGGTTTCCACCTGAA 4858
Db 4801 CCACCTCTGGTTAAAGGAATCAACAGCACAACCAATCCATCTTAAGTGGTTTCCACCTGAA 4860
Qy 4859 GAATCAATGGAACCTCTCTATATATAGCTGGAAGAGAGAGTCACTCTTACAGCT 4918
Db 4861 GAATCAATGGAACCTCTCTATATATAGCTGGAAGAGAGAGTCACTCTTACAGCT 4920
Qy 4919 CTGATGACCAAGATGATGAAAGAAATCCGTTTCAATAGGAAATGGGTATTTGAAATTTCC 4978
Db 4921 CTGATGACCAAGATGATGAAAGAAATCCGTTTCAATAGGAAATGGGTATTTGAAATTTCC 4980
Qy 4979 AGCTCCACTCACCCAGTCAATACAGACTTCACTGGTAACTGTTGACATTTGCTTTAT 5038
Db 4981 AGCTCCACTCACCCAGTCAATACAGACTTCACTGGTAACTGTTGACATTTGCTTTAT 5040
Qy 5039 TAGGAGACAGAGCTTCCAAATATTTTCTATATTTTCAATATCCCTTTTCAATATGAAATTT 5098
Db 5041 TAGGAGACAGAGCTTCCAAATATTTTCTATATTTTCAATATCCCTTTTCAATATGAAATTT 5100
Qy 5099 TATTTATACCTACTTACAGAAATATCTAATTCAGCCCTTTGATAGCTTTTCCCTGATTTT 5158
Db 5101 TATTTATACCTACTTACAGAAATATCTAATTCAGCCCTTTTGTAGCTTTTCCCTGATTTT 5160
Qy 5159 CAGCATGCTCATCTTTTGAATTTCTGGGAAATAGTCAAGTCAAGTGAAGGAAAGAA 5218
Db 5161 CAGCATGCTCATCTTTTGAATTTCTGGGAAATAGTCAAGTCAAGTGAAGGAAAGAA 5220
Qy 5219 AATAAAGATGAAGTGAAGAGCAGCCCTTATTTGGATCAAGTATGCTTTGTATTTGT 5278
Db 5221 AATAAAGATGAAGTGAAGAGCAGCCCTTATTTGGATCAAGTATGCTTTGTATTTGT 5280
Qy 5279 CTTTTTGTGAAGTATGTGCCAGACATGTTTCTGAAATTAATTTATCACTGTGTTCTGGA 5338
Db 5281 CTTTTTGTGAAGTATGTGCCAGACATGTTTCTGAAATTAATTTATTTCACTGTGTTCTGGA 5340
Qy 5339 GCAAATGAGTTTGCAGAAATGCCCTCATGCTATTTGGAGATTCTCAGTATGACCCCTTAC 5398
Db 5341 GCAAATGAGTTTGCAGAAATGCCCTCATGCTATTTGGAGATTCTCAGTATGACCCCTTAC 5400

1019 ATTCACTTAGTGTGAGGTGATCAGACAAAATCAGCTTCTTTATCAATGCGGTGGAG 1078
1021 ATTCACTTAGTGTGAGGTGATCAGACAAAATCAGCTTCTTTATCAATGCGGTGGAG 1080
1079 AAGGATCATACACCTTTCAATGCAAGAACTCTAAGTGGTTCAATACAGATTTGCGATCT 1138
1081 AAGGATCATACACCTTTCAATGCAAGAACTCTAAGTGGTTCAATACAGATTTGCGATCT 1140
1139 GGTACTGTGCAATAGGACAGAGTTTAAATGGTTTAGAGCAGTTGTGCGAAGAAATGCAA 1198
1141 GGTACTGTGCAATAGGACAGAGTTTAAATGGTTTAGAGCAGTTGTGCGAAGAAATGCAA 1200
1199 GATTTTCGATTAATACCAAGTGGACCTTACAAACAGAGAGTTCTGGAAGTCTTCTGGA 1258
1201 GATTTTCGATTAATACCAAGTGGACCTTACAAACAGAGAGTTCTGGAAGTCTTCTGGA 1260
1259 GATCTTCTCAGATGTCATGCCCAATCACAATGCGGTGCGGTGCGAGCCACCGCGGGTC 1318
1261 GATCTTCTCAGATGTCATGCCCAATCACAATGCGGTGCGGTGCGAGCCACCGCGGGTC 1320
1319 CACCTTTTGGCACAGCGGTACTGCAATCTTAATGATGACGAGACACAGCTGATTAATAGA 1378
1321 CACCTTTTGGCACAGCGGTACTGCAATCTTAATGATGACGAGACACAGCTGATTAATAGA 1380
1379 GTGTACCGTTGAATCTGAGGCCCATCTCTCTCTTTTGTCAATGATTAATGATGTTGTT 1438
1381 GTGTACCGTTGAATCTGAGGCCCATCTCTCTCTTTTGTCAATGATTAATGATGTTGTT 1440
1439 ACTTCATGGTTTCAAAATGTTTACAAATTAACATTAACAGCTTAATCAAGGAGTACTATT 1498
1441 ACTTCATGGTTTCAAAATGTTTACAAATTAACATTAACAGCTTAATCAAGGAGTACTATT 1500
1499 TCAGTTGATTTGGAAAATGGACAGTATCAGGTGTTTATATATATCAATCAGTTCTTTAGT 1558
1501 TCAGTTGATTTGGAAAATGGACAGTATCAGGTGTTTATATATATCAATCAGTTCTTTAGT 1560
1559 CACAACCAACCGAATAGGATTAAGAAAGAAAGAAATAGTTTATGATGGAGGAC 1618
1561 CACAACCAACCGAATAGGATTAAGAAAGAAAGAAATAGTTTATGATGGAGGAC 1620
1619 TGGCAATATTTGCCAGGAAATGTTGGTCTTTTGAATGAAACAAATGGAGATTTGGA 1678
1621 TGGCAATATTTGCCAGGAAATGTTGGTCTTTTGAATGAAACAAATGGAGATTTGGA 1680
1679 AAACCTGATTTCTGCACTGCTTCAAGCTTTTCAATTTTACTTCCATATTTCCGCTGGCAAT 1738
1681 AAACCTGATTTCTGCACTGCTTCAAGCTTTTCAATTTTACTTCCATATTTCCGCTGGCAAT 1740
1739 GTCAATTTAGCATCTGACACCTGGACCAAAATTAATGCTCTGGATACAATACTTTCTAT 1798
1741 GTCAATTTAGCATCTGACACCTGGACCAAAATTAATGCTCTGGATACAATACTTTCTAT 1800
1799 AATACCCCATCTTTCAAGAGTCCGTAAGCCACGCAAAATAGGTTTCAATTTCAATGGG 1858
1801 AATACCCCATCTTTCAAGAGTCCGTAAGCCACGCAAAATAGGTTTCAATTTCAATGGG 1860
1859 CAGTACTATAAATCAGACTGCTGTTAACTTACCTCAGACACAGATATTAATGCAAGTGA 1918
1861 CAGTACTATAAATCAGACTGCTGTTAACTTACCTCAGACACAGATATTAATGCAAGTGA 1920
1919 ATCAACATTTAGTGGAGATGTCAGTGCCATGCTATGCTGATGCTGCAACCAAGC 1978
1921 ATCAACATTTAGTGGAGATGTCAGTGCCATGCTATGCTGATGCTGCAACCAAGC 1980
1979 CAGCCATATAGATGCTCTGCTCCAGAGAGCTTCACTGAAGGACTTCAATTTGATGCTG 2038
1981 CAGCCATATAGATGCTCTGCTCCAGAGAGCTTCACTGAAGGACTTCAATTTGATGCTG 2040
2039 TGCTTGGCTCTTTTATAATGCAAGGCTTTTCGCCCAGAGTGAATCAAGTTTACGCTTTCAAT 2098
2041 TGCTTGGCTCTTTTATAATGCAAGGCTTTTCGCCCAGAGTGAATCAAGTTTACGCTTTCAAT 2100
2099 TGTAAACCTTTGTCAATGCAACAGGCAATTCCAAAGCTGCCATTAACAATCTCTGTAGAC 2158

2101 TGTAAACCTTTGTCAATGCAACAGGCAATTCCAAAGCTGCCATTAACAATCTCTGTAGAC 2160
2159 CCATTTCCCTTTTGTAGACATTTAGAGGGGAGAGAGTTTGTGATGATTTGTGAGCATAA 2218
2161 CCATTTCCCTTTTGTAGACATTTAGAGGGGAGAGAGTTTGTGATGATTTGTGAGCATAA 2220
2219 ACTACAGGAAGAACTGTGAGCTGTGCAAGGATTAATTTTCCGACAAAGTGTGTCAGAT 2278
2221 ACTACAGGAAGAACTGTGAGCTGTGCAAGGATTAATTTTCCGACAAAGTGTGTCAGAT 2280
2279 CTTTCGGCCATAGATTTTGTGCAACCTCTGTGATCTGTGATACAGTTTGGCATAGAAATGT 2338
2281 CTTTCGGCCATAGATTTTGTGCAACCTCTGTGATCTGTGATACAGTTTGGCATAGAAATGT 2340
2339 AGCATTTCTTTGTGATCAGATTTGGAGACAGTGTAAATTTGTAAGAGACAGTGTCTGGCAG 2398
2341 AGCATTTCTTTGTGATCAGATTTGGAGACAGTGTAAATTTGTAAGAGACAGTGTCTGGCAG 2400
2399 CAGTGCATCAGTGCAGAAATGGATTTCTAATCTTACAAAGAGTTGATCTCTGATGCTGC 2458
2401 CAGTGCATCAGTGCAGAAATGGATTTCTAATCTTACAAAGAGTTGATCTCTGATGCTGC 2460
2459 AGTCCCTGTAACTGCAATACCTCTGGGACAGTGGATGGAGATTAATCTGTCAACCAAT 2518
2461 AGTCCCTGTAACTGCAATACCTCTGGGACAGTGGATGGAGATTAATCTGTCAACCAAT 2520
2519 TCAGGCGAGTGCAGTGCAGAAAGCAATCTTATTTGGGCTTTAGGTGTGATCAATTTT 2578
2521 TCAGGCGAGTGCAGTGCAGAAAGCAATCTTATTTGGGCTTTAGGTGTGATCAATTTT 2580
2579 GGAATTAATTTTCCGAAAGCTTTTAAATGATGTTGAGTGTGAGCCCTGCGAGTGAACCTC 2638
2581 GGAATTAATTTTCCGAAAGCTTTTAAATGATGTTGAGTGTGAGCCCTGCGAGTGAACCTC 2640
2639 CATGCTCAGTGAACAAATTTCTGCAATCTCTGCGGACAGTGTGAGTGCAGAAAGAA 2698
2641 CATGCTCAGTGAACAAATTTCTGCAATCTCTGCGGACAGTGTGAGTGCAGAAAGAA 2700
2699 GCCAAGGACTTCAAGTGTGACACCTGCAAGGAGACAGTGCAGTGAATGCTTAAG 2758
2701 GCCAAGGACTTCAAGTGTGACACCTGCAAGGAGACAGTGCAGTGAATGCTTAAG 2760
2759 TGTAAAGGCTCTGATGTCACACAGCTGGATCCCTCCCTGGGACTGTCTGTAATGCTTAAG 2818
2761 TGTAAAGGCTCTGATGTCACACAGCTGGATCCCTCCCTGGGACTGTCTGTAATGCTTAAG 2820
2819 ACAGGCGAGTGCATCTGCAAGCCCAATGTTCAAGGAGACAGTGCAGTGAATGCTTAAG 2878
2821 ACAGGCGAGTGCATCTGCAAGCCCAATGTTCAAGGAGACAGTGCAGTGAATGCTTAAG 2880
2879 GGAAATCTTACCTACGCGCAAAATTAATTTTCTGCTGTGCTGCTGCAACCTGTAAG 2938
2881 GGAAATCTTACCTACGCGCAAAATTAATTTTCTGCTGTGCTGCTGCAACCTGTAAG 2940
2939 ACTGGGCAATTAATGCTCTCTGCTGTGTAACAAATCAACAGGACAGTGTCTGTAAG 2998
2941 ACTGGGCAATTAATGCTCTCTGCTGTGTAACAAATCAACAGGACAGTGTCTGTAAG 3000
2999 TTAGGGGTAAACAGGCTCTTCTGCTGTAAATCAGTGTGAGCTTCAAGGACATTTGACCAAT 3058
3001 TTAGGGGTAAACAGGCTCTTCTGCTGTAAATCAGTGTGAGCTTCAAGGACATTTGACCAAT 3060
3059 GACATTTTCAACACTGCGCAGATGTTGATGATTTCTTGGGACATTTACCTGGGACC 3118
3061 GACATTTTCAACACTGCGCAGATGTTGATGATTTCTTGGGACATTTACCTGGGACC 3120
3119 ATTTGTGACCCCAATCAGTGGCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3178
3121 ATTTGTGACCCCAATCAGTGGCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3180
3179 AATCAGTGTCAACAGGCTTTTATATTTTCTCCAGGCAATGCTGCTGCTGCTGCTGCTGCTGCT 3238

Db 4360 CCAGCAGAAAACCTTTTACAAGAGAGAGAAAATTATAGGCTACAGATCAGCATGGTTTCTGAA 4419

QY 4598 CAATCACTCAACAGCTATTCCAGGCGTTTTCACAGCTGTGACACTGCTAAATCC 4657

Db 4420 CATTTCCCTCTACAGATGTTTCAGTCATGTGCTCAAGATGGTCAATTTTCTAGTCC 4479

QY 4658 CAAGAACTATCTTACACTGTAGAGGACTGAAACCTTATAGGATATATAGTATTTACTATT 4717

Db 4480 CAAGACCACTTACATCTGTAAGGAGCTCGAGCCTTATAGGACCTACAGCTTACAGTG 4539

QY 4718 ACTCTCTCAATTCAGTTGGTTGTGTGACAGCTGCTTCGGAGCGAGCAAACTTTAGCA 4777

Db 4540 TCCCTCTCGATTTCGGTGGCTGTGTAAACAGTGTCTTGGGATCATGAGCAAACTTAGCA 4599

QY 4778 GCAG 4781

Db 4600 GCAG 4603

RESULT 3

AY077844

LOCUS 6146 bp mRNA linear ROD 12-AUG-2002

DEFINITION Rattus norvegicus usherin (Ush2a) mRNA, complete cds.

ACCESSION AY077844

VERSION AY077844.1 GI:22212215

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 6146)

AUTHORS Huang, D., Rudy, J.D., Uzvolgyi, E., Davis, J.R., Talmadge, C.B., Pretto, D., Weston, M.D., Lehman, J.E., Zhou, M., Seemayer, T.A., Ahmad, I., Kimberling, W.J., and Sumegi, J.

TITLE Identification of the Mouse and Rat Orthologs of the Gene Mutated in Usher Syndrome Type IIA and the Cellular Source of USH2A mRNA in Retina, a Target Tissue of the Disease

JOURNAL Genomics 80 (2), 195-203 (2002)

MEDLINE 22150873

PUBMED 12160733

REFERENCE 2 (bases 1 to 6146)

AUTHORS Sumegi, J., Huang, D. and Davis, J.R.

TITLE Direct Submission

JOURNAL Submitted (31-JAN-2002) Center for Human Molecular Genetics, University of Nebraska Medical Center, 98545; Nebraska Medical Center, Omaha, NE 68198-5454, USA

FEATURES

source

1. .6146

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

/chromosome="13"

/map="between d13rat49 and d13rat76"

1. .6146

/gene="Ush2a"

183..4721

/gene="Ush2a"

/note="USH2a; extracellular matrix protein; similar Homo sapiens Usher syndrome type IIA protein"

/codon_start=1

/product="usherin"

/protein_id="AA178289.1"

/db_xref="GI:22212216"

/translation="MYIALSSGFLGQAIKTSILAYLSVLLAASQGFPRLENVGA FKVSIVPGHATGCGPSTFCFAAABHAQLCAERLCQDCPYRASPPYALIEGL KSCIPADGDLHPYRSNSTSIFGSHKNCPSLOAPRLAABFTLAVLWKPERGTMV LKTDAGQIVFKVTLSERETMEYRTVNGLOPIKVTMPGRILMKKWIHTVOVHETE VSSFDVGLSEENSTAFDTRTDRSINDSVPSVLICQSLNGSELVGRMQDRFLNLSL TNREILFSGDLPHLHIOQSHCRCPGSHPRVPSVQQICIPNGVEDTIOHRSVRLNPE APLPSFINDDDVATVSWFTDITQLAQGVASIDLENGQVQVQFIPTFRSSPQVPA MRIOKKADKSLWEDQVFAFNCVGMKNNGDLENPNVNCLOPPEIPFISHGNVTF DLTSGQKHPGDDYFYNSSLLQEFWTATQIRLYFRGLFYPAWHTVDSRHRVYAVEI

gene

CDS

TTIGRCOCHGHAECTDRTRPYRCLSPHSPTEGPGCGRCSPLYNDKPPRSNKNVHAP
NCKPQCHGHASCHYDASMDPPPLEYNRGGGVCDDOHHHTGNCSCQDYFYPPI
GADPADPEVCKHCNCRDGTNRGSLLDVGDQCDKRRVSRGRCPFRCHIFIGLQAL
DPDCRPDCNPNFSGVDGDDITCHNSGQCSKANVIGLRCDRCRSPFPLRSLNADGC
EPCHNLAETSVNQLDPLSGQCKBAKGLRCDVCFNRYGLPWSACEVCDNCRAGT
OAGTVDAETGCQCKPSVGRRCSECKEYFNLRQNDLSHLCPCNCKTGTNGSL
CDKSTGQCPCKLGVTLGRCHQCKPHRNLWMDNPQCOACDCDLSGLTGLSGMCDVSG
QCLCLPHROBERRCVQCGCYSSPSNATGCLPCLCHTVAIKNCLNSVGHYCPDPS
TGLSHQCDQDRYFRPBLRCPCHCHVAGASNGTCDATGCGFCKPFTVSGKCDT
CVGASHQDUNLLACSTPQQPRGRVSSAINLSPFPFPFNAHMLTYTLFRD
DSEIYTTDDHPHYTTFYPLTSLSPHTAYSYIETSNVSHSTRIPVCTKPESEG
HLNETHIIPVADSITLWTLGSLNHSPIEKYVLSCTPDHTPEVCSYEGPETSATIR
NLVPTQVCFSGQCTNGSLSSPITVTAQAPQROBPPTWKISPTLKVESVRS
VDNSGLVIRVELYMKRWPSTESLAVFESHGWFHSPASPSANQENLVLDQVSVTL
GLDPHTEYAFRVLAVNAGSVSSNWSASERTGESAPVMAAPSVPLSYLSVSEKPK
ABNTRGRIIGIKLSMVBERSPQRDVPVMSKLVHFAESQDSYIVQRLKPYTYSFT
VSLCASVGCVTSLGEGGCTLTAGKNVLTNTKCTHVGNGYYSRAVLWASYSLSFTAP
SSDLDVVYLESFQKQNH"

BASE COUNT 1553 a 1520 c 1425 g 1634 t 14 others

ORIGIN

Query Match 42.4%; Score 2683.6; DB 10; Length 6146;
Best Local Similarity 75.1%; Pred. No. 0;
Matches 3440; Conservative 2; Mismatches 1106; Indels 34; Gaps 6;

QY 207 GCCATTGCTAAGAACGTCGTTGGTATTACCTTACTCTGAGAACGTCGTCAGTTTCCAG 266

Db 3 GTCTCTCTAAGAACATCTTACTCTGTCTTCCCTGTGAAAGTGTCTGGGCTCTCCA 62

QY 267 AAAATGGAGTATCGCAACATCATCTTAAAGTACCTTCAAGATGTTGCTGGCAAGTGG 326

Db 63 GAGATGGAGATCTCTGTATTCAAGTGCACGCTCGATCCAGTGTTCATAGTCACAG 122

QY 327 CGTGGCTCTGATTTATTATTAGAAATGCTTTATCAGAGAGAGATGCTTTT-TGTAAAC 385

Db 123 AGATCCCTGGACAGTAGTTTAGGAAAGCTTTTCATCAGGTGAAGATATCTTTTCTGTCATC 182

QY 386 ATGAATTCGCCAGTCTTTTCATTCGGCTCTGGCTCTTGTGTTTCAGTCAATGAATGTTG 445

Db 183 ATGATTACCTGGCTCTCTCA-----TCCGGCTTCTGGGTGAGCCATTAAACATCG 236

QY 446 ATCTTTGCTATTGTTTCAATATCTGTATCTGAGTCACGAGGTCTTTTCCCAAGGCTG 505

Db 237 ATCTTGTCTATTGTTGGCTCAGTGTGTGCTGCTCGCAGGGGCTTTTCCCAAGGCTG 296

QY 506 GAGAACGTGGAGCTTTCAAGAAAGTTTCCATCGTGCACACCCAGAGAGTAGTGTGAGCTC 565

Db 297 GAGAACGTGGGTGCTTTCAAGAGGTATCCATTTGTGCAAGCCACGCCACGCTGGATAC 356

QY 566 CCAGACCGAAGCAGCTTTTGTGTCAGAGCTCTGCTGCTGCTGAAAGTATTTCAGTCTGTACC 625

Db 357 CCAGCCCCAGACCTTTTGTGCTAGCGCGCTGGCAGCTGAGCATGGCAGTTGTGTGC 416

QY 626 CAGCGGTTTGTATTTCAGGATGTCCTATACAGATCTTTCACACCTTACCTACACTGCCCTT 685

Db 417 GAGAGGCTGTGTATCCAGGAGTCTCCCTTACCGCTCTGCTCTCCYCCCTATACCGCC-- 474

QY 686 TTCTCAGACGCGCTCAGTAGCTCATACACACAGAGAAGATGATCTGCATCTTAACGCC 745

Db 475 -TCCTAGAGGCGCTCAGGAGTGTGATTCCTCAGACACCGGTGATCTCCACCTTACTCC 533

QY 746 CATAGCAATTCGCAAGTATTATTGGAATCAGAGAGCTGCTTTTCTTCTCTCTCTCT 805

Db 534 AGAGCAACTCCAGAGCTTCATATTGGAAGTCAAGAAGTCTGCTTCTCTCTACAG 590

QY 806 TCTCCAAAGTGATGGCATCAATTATCTTACCTTACCTGTATGGCTGAAACCTGAGCAACAAGT 865

Db 591 GCTCCGAGCTGGCAGCAGATTTTACCTTAGCCGTGTGGTTGAAGCCCGAGGAGGCAGC 650

QY 866 GTAATGTGTGTTATAGAAAAGACGTAGTGGCAGATGTGTGTTCAAACTTACAATATCT 925

Db 651 ACTATGTGTGTGTAGAAAAGACAGCGGATGGGAGATGTGTGTTCAAAAGTTTACAATATCT 710

Db	2871	GACTGTGACTCGTTTGGGGACATTACTTGTGGGAAGCAATGTGTGACCCCGTCACTGGCCAGTCG	2933
Qy	3146	CTGTGTGTGCTCTAATTCGTCAAGGAAGAGTGTAAATCAGTGTCAACACAGGTTTTTATATT	3205
Db	2931	CTATGCCCTACCTCATCGTCAAGAAGAAGGTGTGTACAGTGTCAACAGGGTGTACAGT	2990
Qy	3206	TCTCCAGGCAATGCACTGGCTGCTGCCATGCTCATGCCATCAACCTGCGGCAGTTAAT	3265
Db	2991	TCTCCACAGCAATGCACTGGCTGCTGCCATGCTTATGCCACACAGTTGCTACCAAGAAAT	3050
Qy	3266	CACATCTGTAATAGCTGACTGGTCACTGGTGTGTTTGCCAAAGTGTTCATTCGTGGGCAA	3325
Db	3051	TGCATCTGTAATAGTGTACGGGACATTGCTATTGGCCCTGACCCCTCAACCACTGGGCTG	3110
Qy	3326	CGTTGTGACCAATGCAAGACCAATTACTTTGGATTTTCATCTCAGACTGGAAGATGTCA	3385
Db	3111	TCATGGCACCACTGCCAAGATAGTACTTTAGATTTCATCCCTGACTGGAAGATGCAG	3170
Qy	3386	CCTTGTAAATTGTCATCTCTCAGGAGCCTTGAATGAACCTGTCTACTTGGTTCAGAGCCAG	3445
Db	3171	CCTTGTCACTGTCTATGTAGCAGGAGCCTCGAATGGAACCTTGTGATGCAGTCAAGGCCAG	3230
Qy	3446	TGTTTCTGTAAACAATTTGTCACTGCGCTCAAGTGTGATGCTTGTGTTCCAGTGCAGC	3505
Db	3231	TGTTTCTGCAAGAGATTGTCACTGGCTCAAAATGTGACATTTGTGTTCCTGGTGCAGC	3290
Qy	3506	CACTTGATGTCAACAATCTATTGGGTTGCAGCAAAACTCCATTTCCAGCAACCTCCGCCC	3565
Db	3291	CACTTGACGTCAACAATCTATTAGCTTGTTCAAAATCTCATCTCAGCAACCTCACCC	3350
Qy	3566	AGAGGACAAGTTCAAAGTTCCTTCGTATCAATCTCTCTGGAGTCACCTGATTTCTCCA	3625
Db	3351	AGAGGACGGGTTCAAAGCTCTCTCGCCATCAATCTCTCTCGGAGTCCACTGATTTCCC	3410
Qy	3626	AATGCCCACTGGCTTACTTACAGTTTACTCAGGGATGGTTTGAATCTACACAACAGAG	3685
Db	3411	AATGCACTCTGGCTTACATACACTCTGTTCAGGGATGATTTCCGAGATCTACACACTGAC	3470
Qy	3686	GATCAATPACCATACATACATTTCAATACCTCTTAGACACAGACCTGTTTACCATATACCAA	3745
Db	3471	GATCAGACCCCTTACTATATCTCAGTACTCTCTTGACACTAGTCTGTCAACCGCACAGCA	3530
Qy	3746	TATTCTTATTAATGTAGACACCAATATGTGCATGGTTTCAACAAGGAGTGTAGCTGTCACT	3805
Db	3531	TATTTATTAATACATAGAGACCTCCAAATGTGCACAGTTTCAACAAGGAGCATTCCTGTGATT	3590
Qy	3806	TACAGAACAACCCAGGGTCCAGAGGGGAAACTTGTACTTTAGTTATATCATTCCTTATT	3865
Db	3591	TACAAGACAAGCCAGAGGGCTCAGAAGGCCACTTGAACCTTGACTACATCATTCCTGTG	3650
Qy	3866	GGCTCAGACTCTGTGACACTTACTCTGACACAACACTCTCAAAATCAATCTGGTCCCATAGAG	3925
Db	3651	GCCTCAGACTCCATTACACTTGTCTGACTGGACTCTCCAAACCACTCTGGTCTCTATAGAG	3710
Qy	3926	AAATATATTTTGTCTGTGCCCTTTTGGCTGGTGGTCAGGCATGTGTTTCTTACGAAGGT	3985
Db	3711	AAGTATGTTTTATCATGCACACCTGTGTGGAACAACACTGAGCCCTGTGTCTCTTATGAAGT	3770
Qy	3986	CATGAACCTCAGCTACCATCTGGAATCTGGTTCATTTTGGCAAGTACGATATTTTCTGTA	4045
Db	3771	CCAGAGACCTCAGCTACCATCAGGAATCTGGTTCATTTCCACCCAGTACTGTTTCTCTGTC	3830
Qy	4046	CAGCGGTACTAGCGGGGCTGTTTACACAGCTTGGCCCTATACGTGACCAAGGCCAG	4105
Db	3831	CAGGCTGTACTAATGGGAGCTGTTTATATAGCTCACCTTATACAGTAACCAAGGCCAA	3890
Qy	4106	GCCCTCTCCCAAGACTTAAGTCCACCTTAAGATGCAAAAAATCAGTTCTACAGAACTTCAT	4165
Db	3891	GCGCTCTCCCAAGGACGAGGCCACTTACAGTATGGAATATCAGTCCACAGAACTCAA	3950
Qy	4166	GTAGATGGTCTCCACAGCGGAACCTAAATGGAATAATTAAGATGAATGAATATCATG	4225
Db	3951	GTAGAATGGTCTCGGCCAGTGGATTCAATATGGTGTATATTCAGGTATGATGATATCATG	4010

Qy	4226	AGAAGACTGAGATCTACTATAAAGAAACCA	CATCTGAGGAAAGTCGAGT	TTTTCAGACGAGT	4285
Db	4011	AAGAGATGGCGGTCCA	-----CAGAGGAGAGTCTCGTGT	TTGAGAGCCAT	4055
Qy	4286	GGTTGGCTCAGTCCCTCATTTCA	TTTGTAGAAATGGGCCAA-----	TGAAATGCAATTA	4339
Db	4056	GGTTGGTTCATTTCTCACCAGCCTC	ACCGTCAGCCCAATCAGAGT	GAGAAATGCTTCAG	4115
Qy	4340	CCTCCTCAAAACAAATGACCAACCAT	CACCTGGCTTGGAGCCCATACACCAAGT	ATGAGTTTCAGA	4399
Db	4116	GATCCCAAGTCAGCACAGTCTCTCT	CTGGCTTGGACCCACACACAGAGT	ATGCAATTTAGG	4175
Qy	4400	GTCTTAGCTGTGAATATGAGTGGT	TGGAAGTGTCTTCTGCTCGGTCTC	CAGAAAGACGGGA	4459
Db	4176	GTGTTGGCTGTGAATATGAGTGGC	AGTGTGTCATCTGCTCGGTCTCAGAAAGAC	CAGGA	4235
Qy	4460	GAATCAGACACTGTATTTCAATGAT	CCCTCTTCAGTCTTTCCCTCTCTCTCGT	ACTCTCTC	4519
Db	4236	GAATCAGACACTGTGTTCATGGT	CTCTCTTCAGTCTCTCCACTCT	CACACTACTCCCTC	4295
Qy	4520	AATATCTCTCTGGGGAAGCCAGCAG	ATAATGTTTACAAGAGGAAAAAGTT	TGGGGTATGAC	4579
Db	4296	AGTGTCTCTCTGGGGAAGCCAGCAG	AAAACTTTACAAGAGGAGAGATTA	TAGGGTACAG	4355
Qy	4580	ATCAATATGCTTTCTTGAAACAAT	CCTCAACAGTCTATTTCCATGGCGTTT	TTCACAGCTG	4639
Db	4356	ATCAGCATGGTTTCTGAAAGTTT	CCCCCTCAACGAGATGTTTCCAGT	CATGTGTTCAAAGCTG	4415
Qy	4640	TTGCACACTGCTAAATCCCAAGAA	CTATCTTACATGTAGAGGAGCTG	AAACCTTATAGG	4699
Db	4416	GTGCATTTTGTCTGAGTCCCAAG	ACCAAGTCTTACATCGTTTCAGAGAC	TCAAGCCTTATAGG	4475
Qy	4700	ATATATGAGTTTACTATTACTCT	CTGCAATTCAGTTGGTGTGTGAC	CAAGTGTCTCGGGA	4759
Db	4476	ACCTACAGTTTACTGTCTCCCT	CTGCGCTTCGTGGGCTGTGTAC	CAAGTGTCTTTGGGA	4535
Qy	4760	GCAGGACAAACTTTACAGAG	AG		4781
Db	4536	GAAGGACAAACGCTAACAG	AG		4557

```

RESULT 4
HSUSH2A17
LOCUS
DEFINITION
    HSUSH2A17      2652 bp      DNA      linear      PRI 14-NOV-2000
    Homo sapiens Usher syndrome type IIA protein gene, exon 21 and
    complete cds.
ACCESSION
    AF091889
    AF091889.1      GI:8515085
VERSION
    1
KEYWORDS
    17 of 17
SEGMENT
    Homo sapiens
SOURCE
    Homo sapiens (human)
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
    1 (bases 1 to 2652)
        Weston,M.D., Rudy,J.D., Fugita,S., Yao,S.-P., Usami,S., Cremers,C.,
        Greenburg,J., Ramesar,R., Martini,A., Moiler,C., Smith,R.J.,
        Sumegi,J. and Kimberling,W.J.
        Genomic structure and identification of novel mutations in usherin,
        the gene responsible for Usher syndrome type Iia
        Am. J. Hum. Genet. 66 (4), 1199-1210 (2000)
        20206315
        MEDLINE
        PUBMED
        10729113
AUTHORS
    2 (bases 1 to 2652)
        Weston,M.D., Rudy,J.D., Fugita,T., Sumegi,J. and Kimberling,W.J.
        Direct Submission
TITLE
    Submitted (14-SEP-1998) Genetics, Boys Town National Research
    Hospital, 555 North 30th Street, Omaha, NE 68131, USA
JOURNAL
    Location/Qualifiers
FEATURES
    source
        1..2652
            /organism="Homo sapiens"
            /mol_type="genomic DNA"

```

QY 926 GAGAAAGAGACCAATGTTTATATCGACAGATAAATGTTTGCACCTCCAAATAAAGTA 985
 DB 711 GAGAGAGAGACCAATGTTTATATCGACAGATAAATGTTTGCAGCTCCAAATAAAGTA 770
 QY 986 ATGACACTGGGAGAGATCTCTGGAAGAAATGATTCATCTTAGTGTGAGGTGCATCAG 1045
 DB 771 ATGACACCAGGAGAGATCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 830
 QY 1046 ACATAAATCAGCTCTCTTATCAATGGGTGGAGAGAGATCATACACCTTTCATACCAAGA 1105
 DB 831 ACGGAAGTTCAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 890
 QY 1106 ACTCTAAGGTTCATCAATTCAGATCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1165
 DB 891 ACTCTCGGATTCATCAATTCAGATCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 950
 QY 1166 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1225
 DB 951 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1010
 QY 1226 ACATAAAGAGAGATCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1285
 DB 1011 ACATAAAGAGAGATCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1070
 QY 1286 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1345
 DB 10 1 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1130
 QY 1346 CTAATGATGACAGAGATCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1405
 DB 131 CTAATGATGACAGAGATCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1190
 QY 1406 CTAATGATGACAGAGATCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1465
 DB 1191 CTAATGATGACAGAGATCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1250
 QY 1466 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1525
 DB 1251 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1310
 QY 1526 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1585
 DB 1311 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1370
 QY 1586 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1645
 DB 1371 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1430
 QY 1646 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1705
 DB 1431 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1490
 QY 1706 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1765
 DB 1491 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1550
 QY 1766 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1825
 DB 1551 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1610
 QY 1826 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1885
 DB 1611 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1670
 QY 1886 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1945
 DB 1671 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1730
 QY 1946 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 2005
 DB 1731 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 1790
 QY 2006 AATGCTTTAGACAGCTCTCTGGAAGAAATGATTCATCAGCTGTGAGGTGCATCAG 2065

DB 1791 CACAGCTTCACTGAAGGCCCCAGTGTGTGATGCTCCCTCTTTTATAATGACAGGCT 1850
 QY 2066 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2125
 DB 1851 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 1910
 QY 2126 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2185
 DB 1911 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 1970
 QY 2186 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2245
 DB 1971 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2030
 QY 2246 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2305
 DB 2031 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2090
 QY 2306 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2365
 DB 2091 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2150
 QY 2366 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2425
 DB 2151 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2210
 QY 2426 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2485
 DB 2211 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2270
 QY 2486 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2545
 DB 2271 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2330
 QY 2546 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2605
 DB 2331 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2390
 QY 2606 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2665
 DB 2391 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2450
 QY 2666 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2725
 DB 2451 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2510
 QY 2726 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2785
 DB 2511 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2570
 QY 2786 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2845
 DB 2571 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2630
 QY 2846 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2905
 DB 2631 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2690
 QY 2906 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2965
 DB 2691 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2750
 QY 2966 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 3025
 DB 2751 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2810
 QY 3026 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 3085
 DB 2811 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 2870
 QY 3086 TTCCGCGAAGTGTATCAAGTTTACGCTTTTCAATTTGTAACCTTTGTAACGAAACAGCCAT 3145

QY 6040 AAATATTCATATGTGAACACATGAATGGAATTCGAGCTATATAAAATTTCTATTGTTT 6099
 DB 1906 AAATATTCATATGTGAACACATGAATGGAATTCGAGCTATATAAAATTTCTATTGTTT 1965
 QY 6100 TATTTTCTAAATAGTAATAGTTCTTTTCTATTCAGACTGGCTGCTGATGACCT 6159
 DB 1966 TATTTTCTAAATAGTAATAGTTCTTTTCTATTCAGACTGGCTGCTGATGACCT 2025
 QY 6160 TGGTAATGAATCATGATATATTTCTAACTGAGATATATTGAGATTAATGATGATTAAC 6219
 DB 2036 TGGTAATGAATCATGATATATTTCTAACTGAGATATATTGAGATTAATGATGATTAAC 2085
 QY 6220 ACTCTCTAGTACATCAAAATCAATTCGAGATATTAGAAATTTGAACCTTGAAGCTTAAAA 6279
 DB 2086 ACTCTCTAGTACATCAAAATCAATTCGAGATATTAGAAATTTGAACCTTGAAGCTTAAAA 2145
 QY 6280 TGCTCAACTCTGCTTTATATTTCTAAATGCGAAAAAATTTTTAAAAA 6330
 DB 2146 TGCTCAACTCTGCTTTATATTTCTAAATGCGAAAAAATTTTTAAAAA 2196

RESULT 5
 AC119429/C
 LOCUS 115940 bp DNA linear PRI 19-FEB-2003
 DEFINITION Homo sapiens chromosome 1 clone RP4-723P6, complete sequence.
 AC119429 AC119429 AL1138934
 VERSION HTG
 KEYWORDS AC119429.3 GI:28416171
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 115940)
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and
 Haugen, E.D.
 Direct Submission
 Unpublished
 2 (bases 1 to 115940)
 Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
 Direct Submission
 Submitted (26-APR-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 115940)
 Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
 Direct Submission
 Submitted (11-FEB-2003) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 4 (bases 1 to 115940)
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and
 Haugen, E.D.
 Direct Submission
 Submitted (19-FEB-2003) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 On Feb 19, 2003 this sequence version replaced gi:28301973.
 ----- Genome Center
 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: <http://www.genome.washington.edu>
 Contact: uwgctgs@u.washington.edu
 Drafting Center: SC
 ----- Project Information
 Center project name: chr-1
 Center clone name: RP4-723P6 (sc0827)
 ----- Summary Statistics
 Sequencing vector: plasmid; 75% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 115940 bases at least Q40
 Consensus quality: 115940 bases at least Q30
 Consensus quality: 115940 bases at least Q20

Insert size: 115940; sum-of-contigs
 Quality coverage: 13.5x in Q20 bases; sum-of-contigs

 Overlapping Sequences:
 5': RP11-22M7 (UMGC:sc0603) AC093581, 2474-bp overlap
 3': RP5-1099E6 (UMGC:sc0859) AC138024

Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BglII				HindIII				EcoRI			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9577	9385	1852	1831	3728	3770						
-----	-----	-----	-----	-----	-----						
5671	5768	449	<800	2184	2193						
-----	-----	-----	-----	-----	-----						
10327	10090	512	<800	8065	8167						
-----	-----	-----	-----	-----	-----						
1676	1673	2814	2853	486	<800						
-----	-----	-----	-----	-----	-----						
10319	10090	1247	1234	106	<800						
-----	-----	-----	-----	-----	-----						
494	<800	4695	4783	3581	3414						
-----	-----	-----	-----	-----	-----						
8213	8182	3373	3406	1068	1095						
-----	-----	-----	-----	-----	-----						
1220	1169	3011	3085	1629	1599						
-----	-----	-----	-----	-----	-----						
3450	3448	804	800	313	<800						
-----	-----	-----	-----	-----	-----						
2966	2969	7841	7860	117	<800						
-----	-----	-----	-----	-----	-----						
5992	5956	1345	1337	8123	8167						
-----	-----	-----	-----	-----	-----						
325	<800	4829	4783	5653	5578						
-----	-----	-----	-----	-----	-----						
5999	5956	83	<800	811	821						
-----	-----	-----	-----	-----	-----						
1075	1070	297	<800	3821	3770						
-----	-----	-----	-----	-----	-----						
9479	9385	1092	1095	3464	3596						
-----	-----	-----	-----	-----	-----						
798	815	10	<800	6982	6996						

mRNA	/db_xref="taxon:9606"				
	join(AF091873.1:1.666..848,AF091873.1:1.1522..2210,AF091874.1:1.768..933,AF091875.1:1.343..475,AF091876.1:1.701..764,AF091877.1:1.583..877,AF091878.1:1.749..842,AF091879.1:1.2487..2708,AF091880.1:1.651..781,AF091881.1:1.263..458,AF091882.1:1.411..1052,AF091883.1:1.375..558,AF091884.1:1.561..724,AF091885.1:1.696..854,AF091886.1:1.466..960,AF091887.1:1.487..656,AF091888.1:1.770..914,648..2443)				
	/product="Usher syndrome type IIA protein"				
CDS	join(AF091873.1:1.1726..2210,AF091874.1:1.768..933,AF091875.1:1.343..475,AF091876.1:1.701..764,AF091877.1:1.583..877,AF091878.1:1.749..842,AF091879.1:1.2487..2708,AF091880.1:1.651..781,AF091881.1:1.263..458,AF091882.1:1.411..1052,AF091883.1:1.375..558,AF091884.1:1.561..724,AF091885.1:1.696..854,AF091886.1:1.466..960,AF091887.1:1.487..656,AF091888.1:1.770..914,648..892)				
	/codon_start=1				
	/product="Usher syndrome type IIA protein"				
	/protein_id="AAF75819.1"				
	/db_xref="GI:8515086"				
	/translation="MNCPVLSIGSGFLFOVIEMLIPAYFASISLTESRGLPRLNVGAFKVSIVPTQAVGLPDRSTCHSSAAASISQTCRFQICDQFRRSHPIYITAFSAGLSGCIPTDKNDLPHNHSNASFIFGNHKSFPSPKLMASFTLAVMLKPBQGVWCVLEIKTVGQIVFKLTISEKETMYFTYVNGLPQIPKMTLGLRILVKVNIHLSVQVHQTKISFFINGVEKHQPCNARTLSGISTDPASGTVOIGOSLNGEQFVGRMODRLYQVALNTRILEVPSGDLRLHAQSHCRPCGSHVRPLAORYCIENDAGDTADNRVSRNPEAHLPSFVNDVGVTSVNSVNTITQLNQGTISVDLENGQVQYFIILQFPSPQTEIRIQRKENSIDWEDQYFARNCGAFGMNNDLEKQSVNCLQSLNPTFYSRGNVTFSLTPGPNRYGNNFYNTLSQEFVATQIRPHFGQYVYTTAVNLRIYRYAVNDEITSGRCGHADNCDTSSPYRCLCSQESFTEGLHCDRLPLVNDKPRFGQDQVYAFNFKPCQCNHSHKSYNSVDPFPEHPRGGGVGDDCEHNTTGRNCELDKDYFPRQVADPSAIDVCKPCDCTVGTBNGSILCDQIGGCNCKRVHSGRQCNQONGFYNLOELDPDGSNCINTSGVTDGDIITCHQNSGQCKANVIGLURDCHNFGFKPLKSFNDVGCPCQCNLHGSVKNPCNPHSGQCKEAKGLQCDTCRENFYGLDVTNCKACDDTAGSLPFCNAKTQOCICKPNEVRQCNKLEGNFYLRONNSFLCLPCNCDKTGTINGSLCNKTSQGCCKLGVTVLRNQCEPHRYNLTIDNFQHCQMCDCSLGTPTGICVDPISSGQCLVQPNRQCNQCPQFVISPQNAAGCLPSCHTTGAVNHI CNSITGOCVCQDASITAGORCDQKHIFGDFPQTCRCQPCNCHLSGALNETCHLVYTGOCFCQKPYTGSKDCAVPSASHLDVNNLLGCKSTPQQPPRGQVQSSAINLSWSPDPSFNARMLTYSLLRDGFEIYTTEDQPYSTQYFLTDLLPYTKYISYIETNVHGSVTRSVATYTKPGVPEGNLIIYIPDGSVTLTWTLSNOSGPIEKYIILSCAPLAGQPCQSVYEGHETSATIMNLVPPAKYDFSVQACTSGGCLHSLPITVTTAQAPQRLSPPMOKISSTELHVEWSPPAELANGIIIRYELVYMRRLSTKETTSEERSVFOSSGWLSPHSFVPSANENALKPPQTMVITGLPEYTKYEFVRLVNMAGSVSSAWVSERTGESAPVFMIPSPVPLSSYSLNLSWEKPADNVTYKGVGVGDVNNLSEQSPQSI PNAPSQLLHAKSLSYTVBGLKPYRIYFTITLQNSVGVTSASAGQTLAAAPQLRPLVPLVKGINSTTILHLKWPPEELNGPSPYQLERESSLPALMTMMKIRFIGNGYCKFPSSSTHPVNDPTDKCV"				
exon	648..2443				
	/number=21				
BASE COUNT	840 a	500 g	832 t	1 others	
ORIGIN					
Query Match	24.4%; Score 1546.2; DB 9; Length 2652;				
Best Local Similarity	99.8%; Pred. No. 0;				
Matches 1548; Conservative	0; Mismatches 3; Indels 0; Gaps 0;				
QY	4780	AGCAGCAGCAACTGAGCCACCTCTGGTTTAAAGAAATCAACAGCACCAACATCATCT			4839
Db	646	AGCAGCAGCAACTGAGGCCACCTCTGGTTTAAAGAAATCAACAGCACCAATCATCT			705
QY	4840	TAAGTGGTTTCCACCTGAGAACTGAATGGACCTCTCTATATATATCAGCTGGAAAGAG			4899
Db	706	TAAGTGGTTTCCACCTGAGAACTGAATGGACCTCTCTATATATATCAGCTGGAAAGAG			765
QY	4900	AGAGTCATCTCTACCAGCTCTGATGACCAAGATGAAGAAATCCGTTTCATAGAAA			4959
Db	766	AGAGTCATCTCTACCAGCTCTGATGACCAAGATGAAGAAATCCGTTTCATAGAAA			825

QY	4960	TGGGTATTGTAAATTTCCAGCTCCACTCCAGTCAATACAGACTTCACTGGTAAGTG	5019		
DB	826	TGGGTATTGTAAATTTCCAGCTCCACTCCAGTCAATACAGACTTCACTGGTAAGTG	885		
QY	5020	TGTTTGACATCTCTTATTATAGGACACGAGCTCCAAAATGTTCTATATTTTCATA	5079		
DB	886	TGTTTGACATCTCTTATTATAGGACACGAGCTCCAAAATGTTCTATATTTTCATA	945		
QY	5080	TCCCTTTCAATGAATTTTATTATACCTTACTTAGAGAAATCTAATTCAGGCCCTTGTAT	5139		
DB	946	TCCCTTTCAATGAATTTTATTATACCTTACTTAGAGAAATCTAATTCAGGCCCTTGTAT	1005		
QY	5140	AGCTTTGCTGCTGATTTTTCAGCATGTCATCTTTTGAAGATCTGGGGAAGAAAGTCAG	5199		
DB	1006	AGCTTTGCTGCTGATTTTTCAGCATGTCATCTTTTGAAGATCTGGGGAAGAAAGTCAG	1065		
QY	5200	GTAAGTGAAGGAAAGAAATAAAGATGAAGATGAAGAGACAGCCTTATTTGGATCAA	5259		
DB	1066	GTAAGTGAAGGAAAGAAATAAAGATGAAGATGAAGAGACAGCCTTATTTGGATCAA	1125		
QY	5260	GTATGTGCTTTGTATTGTCTTTTGTGAAGTATGTCCAGGACATGTTCTTGAATAT	5319		
DB	1126	GTATGTGCTTTGTATTGTCTTTTGTGAAGTATGTCCAGGACATGTTCTTGAATAT	1185		
QY	5320	TATTCACCTGTGTTCTTCAGGCAATAGTTTGCAAAATGCCCTCATGCTATTTGGAGATTC	5379		
DB	1186	TATTCACCTGTGTTCTTCAGGCAATAGTTTGCAAAATGCCCTCATGCTATTTGGAGATTC	1245		
QY	5380	TCAGTATGACACCCCGTACTGAACTCCAAAGACATTTGAAGAGCTTATTTCAACTTTG	5439		
DB	1246	TCAGTATGACACCCCGTACTGAACTCCAAAGACATTTGAAGAGCTTATTTCAACTTTG	1305		
QY	5440	CTTAGCTAATCATGCTTAAACAGATATTTGATGATGTTTCTTTTCTTTCTCTTCTGCTG	5499		
DB	1306	CTTAGCTAATCATGCTTAAACAGATATTTGATGATGTTTCTTTTCTTTCTCTTCTGCTG	1365		
QY	5500	TTTCTCTCTCTTTTCTCTGTCGACAACTTAATATCTCATGTTCTATGAAGAAACATTTG	5559		
DB	1366	TTTCTCTCTCTTTTCTCTGTCGACAACTTAATATCTCATGTTCTATGAAGAAACATTTG	1425		
QY	5560	TGGGGAAGAACTAAATCCAGGGAAGAGATACTCTCTTAAGCAGCACTATGTAAGCAA	5619		
DB	1426	TGGGGAAGAACTAAATCCAGGGAAGAGATACTCTCTTAAGCAGCACTATGTAAGCAA	1485		
QY	5620	GTAGGCTCTTGTGTTTGGTCTCAAAATTTAAAGGCACTTAAAGGCACTATGTAAGCAA	5679		
DB	1486	GTAGGCTCTTGTGTTTGGTCTCAAAATTTAAAGGCACTTAAAGGCACTATGTAAGCAA	1545		
QY	5680	ATTTTAATGCAATATTTTAAATAATGAAATCAATGTGAAAGCACTATAAAATATATTC	5739		
DB	1546	ATTTTAATGCAATATTTTAAATAATGAAATCAATGTGAAAGCACTATAAAATATATTC	1605		
QY	5740	AAAAGCTTAAATAAGACAGATTTGAACTCTCTGACAGCACTATCTGCTCTCTGCTT	5799		
DB	1606	AAAAGCTTAAATAAGACAGATTTGAACTCTCTGACAGCACTATCTGCTCTCTGCTT	1665		
QY	5800	ACCTCTCTCTGCTTACTAGTACCGCAATATTTTGAAGTCCCATGACCTCTGCTGACT	5859		
DB	1666	ACCTCTCTCTGCTTACTAGTACCGCAATATTTTGAAGTCCCATGACCTCTGCTGACT	1725		
QY	5860	TACAGCTTCTAATAGCATGATTTCAATATAGCTGTAAATAAATCTCTACTTATGTTACACC	5919		
DB	1726	TACAGCTTCTAATAGCATGATTTCAATATAGCTGTAAATAAATCTCTACTTATGTTACACC	1785		
QY	5920	ATTTTTCATTTTAAATAAATTTTCAAGTATAAGATATATATTTATTTATGTTAACTCA	5979		
DB	1786	ATTTTTCATTTTAAATAAATTTTCAAGTATAAGATATATATTTATTTATGTTAACTCA	1845		
QY	5980	TAAAGATGTTTCAATTAATCATCCATGAGAAAGTCAATTTTGGAGCAATAGCTAGCTTTTA	6039		
DB	1846	TAAAGATGTTTCAATTAATCATCCATGAGAAAGTCAATTTTGGAGCAATAGCTAGCTTTTA	1905		

QY 5740 AAAAGCTTAATAAGACAGATGAACTCTGTACGACCAATCTCTGCTCTACTGCGCTT 5799
 |||||
 Db 85169 AAAAGCTTAATAAGACAGATGAACTCTGTACGACCAATCTCTGCTCTACTGCGCTT 85110
 |||||
 QY 5800 ACCCTCTCTGCGCTTACTAGTACCGCAATATTTTGGAGTCCCAATGACCTCTGTGACT 5859
 |||||
 Db 85109 ACCCTCTCTGCGCTTACTAGTACCGCAATATTTTGGAGTCCCAATGACCTCTGTGACT 85050
 |||||
 QY 5860 TACAGCTTCTAATAGCATGATTTCAATATAGCTGTAAATAAACTCTACTTATGGTACACC 5919
 |||||
 Db 85049 TACAGCTTCTAATAGCATGATTTCAATATAGCTGTAAATAAACTCTACTTATGGTACACC 84990
 |||||
 QY 5920 ATTTTTCCTAATTTTAAAAAATTTTCAAAAGTATGAATATATTTATGTAATCTCA 5979
 |||||
 Db 84989 ATTTTTCCTAATTTTAAAAAATTTTCAAAAGTATGAATATATTTATGTAATCTCA 84930
 |||||
 QY 5980 TAAAGATGTTCAATTAATCATCATGAGAAAGTCAATTTTGGAGCAAAATAGTACTTTA 6039
 |||||
 Db 84929 TAAAGATGTTCAATTAATCATCATGAGAAAGTCAATTTTGGAGCAAAATAGTACTTTA 84870
 |||||
 QY 6040 AATATTTGCATATGTAAGACATGAAATGGAATTCGAGCTATAAAATTTGTATGTTT 6099
 |||||
 Db 84869 AATATTTGCATATGTAAGACATGAAATGGAATTCGAGCTATAAAATTTGTATGTTT 84810
 |||||
 QY 6100 TATTTTACTTAAATAGTAAATAGTTTCTTTTCAATGAGACTGGCTGCTGATGCACCT 6159
 |||||
 Db 84809 TATTTTACTTAAATAGTAAATAGTTTCTTTTCAATGAGACTGGCTGCTGATGCACCT 84750
 |||||
 QY 6160 TGGTAAATGATCATGATTAATTTCTAACTGAGATATATTTGAGATTAATGATGATTA 6219
 |||||
 Db 84749 TGGTAAATGATCATGATTAATTTCTAACTGAGATATATTTGAGATTAATGATGATTA 84690
 |||||
 QY 6220 ACTCTCTCAGTACATCAAAATCATTGCGAGTATTAGAAATGACCATGAGCTAAAAA 6279
 |||||
 Db 84689 ACTCTCTCAGTACATCAAAATCATTGCGAGTATTAGAAATGACCATGAGCTAAAAA 84630
 |||||
 QY 6280 TGCTCAACTCTGCTTTATTTCTTAAATGCGCAAAAAAATTTTAAAAA 6330
 |||||
 Db 84629 TGCTCAACTCTGCTTTATTTCTTAAATGCGCAAAAAAATTTTAAAAA 84579
 |||||

RESULT 6

AC138024/c
 LOCUS AC138024 163815 bp DNA linear PRI 01-MAR-2003
 DEFINITION Homo sapiens chromosome 1 clone RP5-1099E6, complete sequence.
 ACCESSION AC138024 AL38782
 VERSION AC138024.2 GI:28626669
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and
 Haugen, E.D.
 Direct Submission
 Unpublished
 2 (bases 1 to 163815)
 Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
 Direct Submission
 Submitted (10-DEC-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 163815)
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and
 Haugen, E.D.
 Direct Submission
 Submitted (01-MAR-2003) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 On Mar 1, 2003 this sequence version replaced gi:26291485.
 ----- Genome Center
 Center: University of Washington Genome Center

Center Code: UMGC
 Web site: <http://www.genome.washington.edu>
 Contact: uwgctg@u.washington.edu
 Drafting Center: SC

----- Project Information

Center project name: chr-1
 Center clone name: RP5-1099E6 (sc0859)

----- Summary Statistics

Sequencing vector: plasmid; 82% of reads
 Sequencing vector: plasmid; 108752; 18% of reads
 Chemistry: Dye-terminator ET; 20% of reads
 Chemistry: Dye-terminator Big Dye; 80% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 163790 bases at least Q40
 Consensus quality: 163815 bases at least Q30
 Insert size: 163815; sum-of-contigs
 Quality coverage: 8.9x in Q20 bases; sum-of-contigs

----- Overlapping Sequences:

5': RP4-723P6 (UMGC:sc0827) AC119429, 100473-bp overlap
 3': RP11-239122 AL358858, 100-bp overlap

----- Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

EcoRI			HindIII			BglII		
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap
5829	5793	1454	1511	11401	11165			
2184	2180	449	<800	5671	5862			
8065	8081	512	<800	13046	12992			
486	<800	2814	2667	1819	1798			
106	<800	1247	1246	3902	3914			
4368	4483	4695	4872	5005	5109			
1070	1062	3373	3343	3847	3787			
3044	3127	11328	10882	13968	14281			

1023	1016	2363	2230	9032	8951	1020	1016	15331	15830
2764	2870	4285	4490	3184	3134	524	<800	3891	3870
83	<800	1862	1935	4338	4430	1353	1320	583	<800
2856	3012	3533	3343	7186	7196	4589	4483	550	<800
2011	1970	3192	3343	3045	3020	673	<800	3872	3870
4071	3999	3729	3870	4618	4838	3419	3396	1550	1580
15	<800	1293	1246	2827	2794	1529	1479	114	<800
18676	18998	128	<800	29	<800	6982	6935	1989	2039
1216	1201	6192	6130	7295	7196	3464	3396	459	<800
4182	4323	709	<800	7254	7196	3821	3783	2991	2840
3994	3999	3355	3343	5375	5622	811	817	4471	4490
119	<800	3037	3055	5244	5280	1421	1422	1421	1422
2538	2539	4420	4490	2663	2638				
1811	1782	1526	1580	463	<800				
3123	3267	623	<800	4549	4529				
1440	1405	1206	1246	2980	2919				
637	<800	4014	4140	4004	4239				
2972	3012	31	<800	873	896				
66	<800	4933	5103	6196	6116				
2591	2620	943	843	798	823				
229	<800	244	<800	9479	9364				
1358	1320	3156	3055	1075	1071				
8081	8081	660	<800	5999	5862				
13	<800	1591	1668	325	<800				
803	817	12379	12010	5992	5862				
5206	5155	172	<800	2966	2919				
4030	3999	203	<800	3450	3373				
7174	7174	4074	4140	1220	1178				
4168	4154	452	<800	8213	8143				
892	904	1024	1082	494	<800				
9091	8928	1081	1082						
2109	2081	1072	1082						
9521	9349	4339	4490						
2653	2749	2883	2840						
11613	11225	2453	2285						
3359	3396	7821	7671						

Query Match 24.4%; Score 1546.2; DB 9; Length 163815;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1548; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	4780	AGCACGACAACTGAGGCCACCTCTGGTTAAAGNAATCAACAGCACAAATCCATCT	4839
DB	70662	AGCACGACAACTGAGGCCACCTCTGGTTAAAGNAATCAACAGCACAAATCCATCT	70603
QY	4840	TAAGTGGTTCCACCTGAGAACTGAATGGACCTCTCTCTATATATCAGCTGGAAAGGAG	4899
DB	70602	TAGGTGGTTCCACCTGAGAACTGAATGGACCTCTCTCTATATATCAGCTGGAAAGGAG	70543
QY	4900	AGAGTCATCTTACCAGCTCTGATGACCCAGATGATGAAGAAATCCGTTTCATAGGAA	4959
DB	70542	AGAGTCATCTTACCAGCTCTGATGACCCAGATGATGAAGAAATCCGTTTCATAGGAA	70483
QY	4960	TGGGTATTGTAATTTCCAGCTCCACTCACCAGTCAATACAGACTTCACCTGGTAAGTG	5019
DB	70482	TGGGTATTGTAATTTCCAGCTCCACTCACCAGTCAATACAGACTTCACCTGGTAAGTG	70423
QY	5020	TGTTTGACATTCGTTTATTTAGGAGACAGAGCTCCAAAATGTTTCTATATTTTCATA	5079
DB	70422	TGTTTGACATTCGTTTATTTAGGAGACAGAGCTCCAAAATGTTTCTATATTTTCATA	70363
QY	5080	TCCCTTTACATGAATTTTATTTATATACCTTACCTTAGAGAAATCTAATTCAGCCCTTGAT	5139
DB	70362	TCCCTTTACATGAATTTTATTTATATACCTTACCTTAGAGAAATCTAATTCAGCCCTTGAT	70303
QY	5140	AGCTTTTCCTGATGTTTCAGCATGCTCCATCTTTTGAATTTCTGGGAAAAAAGTCAG	5199
DB	70302	AGCTTTTCCTGATGTTTCAGCATGCTCCATCTTTTGAATTTCTGGGAAAAAAGTCAG	70243
QY	5200	GTAAGTGAAGAAAAAGAAAAATAAAGATGAAGATGAAGAGCAGCTTTATTTGGATCAAA	5259
DB	70242	GTAAGTGAAGAAAAAGAAAAATAAAGATGAAGATGAAGAGCAGCTTTATTTGGATCAAA	70183
QY	5260	GTAAGTGTCTTTGATTTCTTTTGTGAAGTATGCGAGACATGTTCTTGAATAT	5319
DB	70182	GTAAGTGTCTTTGATTTCTTTTGTGAAGTATGCGAGACATGTTCTTGAATAT	70123
QY	5320	TATTCACGTGTTCTCTGAGCAATGAGTTTGCAAAATGCCCTCATGCTATTTGGAGATTC	5379
DB	70122	TATTCACGTGTTCTCTGAGCAATGAGTTTGCAAAATGCCCTCATGCTATTTGGAGATTC	70063
QY	5380	TCAGTATGACCCCGTTTACTGAACTCCAAAAGCATTTGAAGAAAGCTTATTTCAACTTTG	5439
DB	70062	TCAGTATGACCCCGTTTACTGAACTCCAAAAGCATTTGAAGAAAGCTTATTTCAACTTTG	70003

AUTHORS Weston, M.D., Eudy, J.D., Fugita, S., Yao, S.-F., Usami, S., Cremers, C., Greenburg, J., Ramesar, R., Martini, A., Moller, C., Smith, R.J., Sumegi, J., and Kimberling, W.J.
TITLE Genomic structure and identification of novel mutations in usherin, the gene responsible for Usher syndrome type IIa
JOURNAL Am. J. Hum. Genet. 66 (4), 1199-1210 (2000)
MEDLINE 20206315
PUBMED 10729113
REFERENCE 2 (bases 1 to 2749)
AUTHORS Weston, M.D., Eudy, J.D., Fugita, T., Sumegi, J., and Kimberling, W.J.
TITLE Direct Submission
JOURNAL Submitted (14-SEP-1998) Genetics, Boys Town National Research Hospital, 555 North 30th Street, Omaha, NE 68131, USA
FEATURES
source
1. .2749
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
666..848
/number=1
1522..2210
/number=2
BASE COUNT 810 a 536 c 561 g 842 t
ORIGIN
Query Match
Best Local Similarity 10.8%; Score 683.4; DB 9; Length 2749;
Matches 701; Conservative 1; Mismatches 12; Indels 1; Gaps 1;
QY 182 GATACAGCAGCTACTCATGCTCTCGCCATTGCTAAGACGCTGGTGGTATTACCTTACT 241
DB 1521 GATACAGCAGCTACTCATGCTCTCGCCATTGCTAAGACGCTGGTGGTATTACCTTACT 1580
QY 242 CTGAGAACCTGTCTGCGAGTTTCAGAAATGAGATATCGCAATCATCACTTAAAGTACCT 301
DB 1581 CTGAGAACCTGTCTGCGAGTTTCAGAAATGAGATATCGCAATCATCACTTAAAGTACCT 1640
QY 302 GCTTCAAGTATTGTCGCAAGTGGCGTGGCTGATATTATTATTAGAAATGCTTTATC 361
DB 1641 GCTTCAAGTATTGTCGCAAGTGGCGTGGCTGATATTATTATTAGAAATGCTTTATC 1700
QY 362 AGGAGGAGAAATGC - TTTTGTGAACATGAATGCCAGTCTTTTCATTGGGCTCTGGCTT 420
DB 1701 AGGAGGAGAAATGCTTTTGTGAACATGAATGCCAGTCTTTTCATTGGGCTCTGGCTT 1760
QY 421 CTGTGTTTCAAGTATGAATGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTT 480
DB 1761 CTGTGTTTCAAGTATGAATGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTT 1820
QY 481 GTCAAGAGTCTTTTCCCAAGCTGGAGAACGTGGAGCTTTCAAGAAAGTTTCCATGCT 540
DB 1821 GTCAAGAGTCTTTTCCCAAGCTGGAGAACGTGGAGCTTTCAAGAAAGTTTCCATGCT 1880
QY 541 GCCAACCCAGCAGTATGTGGATCTCCAGACCGAAGCACTTTTGTTCAGAGCTCTGCTGC 600
DB 1881 GCCAACCCAGCAGTATGTGGATCTCCAGACCGAAGCACTTTTGTTCAGAGCTCTGCTGC 1940
QY 601 TGTGTAAGTATTCAGTCTCTGATCCAGCGGTTTGTATTCAGAGATGCGCCATACAGATC 660
DB 1941 TGTGTAAGTATTCAGTCTCTGATCCAGCGGTTTGTATTCAGAGATGCGCCATACAGATC 2000
QY 661 TTCACACCTTACTACACGCTTCTTCTCAGCAGGCTCTCAGTCTGATCATCACACCGA 720
DB 2001 TTCACACCTTACTACACGCTTCTTCTCAGCAGGCTCTCAGTCTGATCATCACACCGA 2060
QY 721 CAAGATATGATCTGATCTTCAAGCCCATAGCAATCTCGAAGTTTATTTTGGAAATCA 780
DB 2061 CAAGATATGATCTGATCTTCAAGCCCATAGCAATCTCGAAGTTTATTTTGGAAATCA 2120
QY 781 CAAGAGCTGCTTTTCTTCT 840
DB 2121 CAAGAGCTGCTTTTCTTCT 2180

QY 841 ATGGCTGAACCTGAGCAACAGGTGTAATGTTGTTATAGAAAGACGTAGAT 895
DB 2181 ATGGCTGAACCTGAGCAACAGGTGTAATGTTGTTATAGAAAGACGTAGAT 2235
RESULT 9
AL445650/c
LOCUS Human DNA sequence from clone RP11-152K20 on chromosome 1, complete sequence.
DEFINITION AL445650
ACCESSION AL445650
VERSION AL445650.9 GI:11611403
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 153705)
AUTHORS Martin, S.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
REQUESTS: clonerequest@sanger.ac.uk
COMMENT On Dec 9, 2000 this sequence version replaced gi:11602591.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
- EMBL, SW, SWISSPROT, Tr, TREMBL, Wp, WORMPEP, Information from the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-152K20 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-152K20 The true right end of clone RP11-239122 is at 26243 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.
FEATURES
source
1. 153705
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-152K20"
/clone_lib="RPCI-11.1"
3. 481
/note="match: GSS: Em:AQ383997"
368..666
/note="AluSx repeat: matches 1..303 of consensus"
936..986
/note="L2 repeat: matches 2700..2750 of consensus"
947..1184
/note="MIR repeat: matches 32..257 of consensus"
1181..1740
/note="match: GSS: Em:AQ590646"
1185..1838
/note="match: GSS: Em:A2520510"
1201..1394
misc_feature
repeat_region
repeat_region
repeat_region
misc_feature
misc_feature
misc_feature


```
repeat_region /note="match: GSS: Em:AQ503999"
1219. .1340
repeat_region /note="61 copies 2 mer ta 64% conserved"
1229. .1340
repeat_region /note="28 copies 4 mer tata 66% conserved"
2158. .2568
repeat_region /note="MLT1C repeat: matches 54. .466 of consensus"
3375. .3617
repeat_region /note="MER58A repeat: matches 1. .223 of consensus"
4003. .4172
misc_feature /note="L1MC1 repeat: matches 6138. .6306 of consensus"
complement(4777. .5298)
misc_feature /note="match: GSS: Em:AQ525021"
5923. .6252
misc_feature /note="match: GSS: Em:AQ152396"
complement(5934. .6086)
repeat_region /note="match: GSS: Em:AQ186738"
7304. .7337
repeat_region /note="MIR repeat: matches 229. .262 of consensus"
7338. .7648
repeat_region /note="AluSg repeat: matches 1. .301 of consensus"
7649. .7765
repeat_region /note="MIR repeat: matches 111. .230 of consensus"
7784. .8093
repeat_region /note="AluSx repeat: matches 1. .309 of consensus"
8095. .8409
misc_feature /note="AluS1 repeat: matches 1. .304 of consensus"
complement(10099. .10432)
repeat_region /note="match: GSS: Em:AQ529802 Em:AQ530310"
11036. .11169
repeat_region /note="L2 repeat: matches 1170. .1301 of consensus"
11235. .11506
repeat_region /note="L1MC4 repeat: matches 7574. .7834 of consensus"
11507. .11964
repeat_region /note="L1PA4 repeat: matches 5689. .6146 of consensus"
11965. .12330
repeat_region /note="L1MC4 repeat: matches 7172. .7574 of consensus"
12351. .12698
repeat_region /note="L1MC4 repeat: matches 6643. .6996 of consensus"
12870. .13925
repeat_region /note="TIGSRR1 repeat: matches 55. .1118 of consensus"
15429. .15642
repeat_region /note="107 copies 2 mer aa 72% conserved"
15442. .15637
misc_feature /note="49 copies 4 mer gaaa 82% conserved"
16549. .17186
misc_feature /note="match: GSS: Em:AQ239802"
17713. .18197
repeat_region /note="match: GSS: Em:AQ613916"
19478. .19712
misc_feature /note="MER8 repeat: matches 11. .239 of consensus"
complement(19904. .20377)
repeat_region /note="match: GSS: Em:AQ885526"
21383. .21707
repeat_region /note="AluJb repeat: matches 1. .302 of consensus"
21843. .21905
misc_feature /note="L1PA8 repeat: matches 6101. .6163 of consensus"
23594. .24161
misc_feature /note="match: GSS: Em:AQ377092"
23595. .24035
repeat_region /note="match: GSS: Em:AQ507156"
23923. .24218
misc_feature /note="AluY repeat: matches 1. .308 of consensus"
24217. .24225
/note="Sequence from overlapping clone ba239122
(AL358858). Assembly confirmed by restriction digest
data."
repeat_region 24249. .24272
/note="12 copies 2 mer ta 100% conserved"
24274. .24574
repeat_region /note="AluY repeat: matches 2. .294 of consensus"
complement(25249. .25594)
misc_feature /note="match: GSS: Em:AQ551687"

complement(25692. .26243)
/note="match: GSS: Em:AQ561898"
29416. .29473
/note="MIR repeat: matches 88. .146 of consensus"
29658. .29689
/note="MIR repeat: matches 106. .137 of consensus"
29924. .30224
/note="AluY repeat: matches 1. .307 of consensus"
31517. .32047
/note="L1MC/D repeat: matches 5154. .5658 of consensus"
32428. .32724
/note="AluY repeat: matches 1. .297 of consensus"
33358. .33581
/note="MER20 repeat: matches 8. .213 of consensus"
35048. .35677
/note="MLT2D repeat: matches 1. .548 of consensus"
36085. .36235
/note="L1MD repeat: matches 232. .380 of consensus"
36236. .36275
/note="20 copies 2 mer ca 87% conserved"
36239. .36278
/note="10 copies 4 mer acac 87% conserved"
37305. .37373
/note="L1MD repeat: matches 542. .623 of consensus"
37651. .38715
/note="L1MD repeat: matches 980. .2053 of consensus"
39119. .39615
/note="L1M4 repeat: matches 2110. .2596 of consensus"
39808. .40105
/note="L1M4 repeat: matches 2679. .2947 of consensus"
40106. .40535
/note="Tigger2a repeat: matches 2. .431 of consensus"
40536. .40614
/note="L1M4 repeat: matches 2947. .3030 of consensus"
40660. .40740
/note="L1M4 repeat: matches 3016. .3101 of consensus"
40731. .41438
/note="L1MD2 repeat: matches 3750. .4490 of consensus"
41439. .41818
/note="TH21C repeat: matches 1. .371 of consensus"
41819. .42252
/note="L1MD2 repeat: matches 4490. .4914 of consensus"
42326. .42630
/note="AluJb repeat: matches 1. .303 of consensus"
42662. .42862
/note="AluJo repeat: matches 1. .220 of consensus"
42877. .43009
/note="L1MD2 repeat: matches 5017. .5153 of consensus"
43039. .44071
/note="L1MD2 repeat: matches 5290. .6335 of consensus"
44769. .44844
/note="19 copies 4 mer ctic 75% conserved"
46060. .46764
/note="MER44C repeat: matches 1. .728 of consensus"
46827. .48181
/note="L1P4 repeat: matches 4387. .5765 of consensus"
48240. .48346
/note="MER5B repeat: matches 60. .177 of consensus"
48757. .48912
/note="L2 repeat: matches 2168. .2335 of consensus"
49037. .49302

Query Match 10.8%; Score 683.4; DB 9; Length 153705;
Best Local Similarity 98.0%; Pred No 1.8e-140;
Matches 701; Conservative 1; Mismatches 13; Indels 1; Gaps 1;

QY 182 GATACACGACTACTCATGTCTTGCCATTGCTAAGAACGCTGTGGTATTACTTACT 241
DB 60248 GATACACGACTACTCATGTCTTGCCATTGCTAAGAACGCTGTGGTATTACTTACT 60189
QY 242 CTGAGAACGTCTGTCAGTTTCCAGAAATGAGATATCGCAACATCACTTAAAGTACCCT 301
DB 60188 CTGAGAACGTCTGTCAGTTTCCAGAAATGAGATATCGCAACATCACTTAAAGTACCCT 60129
```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

HTG; HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 290708)
McLay, K.
Direct Submission
Submitted (16-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Sep 27, 2002 this sequence version replaced gi:22415846.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj865G23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 0% of reads
Sequencing vector: plasmid; L08752; 99% of reads
Chemistry: Dye-terminator ABI; 0% of reads
Chemistry: Dye-terminator; 21% of reads
Chemistry: Dye-terminator ET-amersham; 9% of reads
Dye-terminator Big Dye; 68% of reads
Consensus quality: 276213 bases at least Q40
Consensus quality: 280417 bases at least Q30
Consensus quality: 282780 bases at least Q20
Insert size: 285208; sum-of-contigs
Insert size: 134879; 2.4% error; agarose-fp
Quality coverage: 6.08x in Q20 bases; sum-of-contigs Quality
coverage: 15.17x in Q20 bases; agarose-fp

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1. 2509: contig of 2509 bp in length
2510: contig of 100 bp
2610: 9943: contig of 7334 bp in length
9944: 10043: gap of 100 bp
10044: 12149: contig of 2106 bp in length
12150: 12249: gap of 100 bp
12250: 15904: contig of 3655 bp in length
15905: 16004: gap of 100 bp
16005: 20365: contig of 4561 bp in length
20366: 20665: gap of 100 bp
20666: 25474: contig of 4809 bp in length
25475: 25574: gap of 100 bp
25575: 27837: contig of 2263 bp in length
27838: 27937: gap of 100 bp
27939: 30564: contig of 2627 bp in length
30565: 30664: gap of 100 bp
30666: 34794: contig of 4130 bp in length
34795: 34894: gap of 100 bp
34895: 52088: contig of 17194 bp in length
52089: 52188: gap of 100 bp
52189: 58992: contig of 6804 bp in length
58993: 59092: gap of 100 bp
59093: 63330: contig of 4238 bp in length
63331: 63430: gap of 100 bp
63431: 65806: contig of 2376 bp in length
65807: 65906: gap of 100 bp
65907: 69790: contig of 3884 bp in length
69791: 69890: gap of 100 bp
69891: 77440: contig of 7550 bp in length
77441: 77540: gap of 100 bp

77541: 84031: contig of 6491 bp in length
84032: 84131: gap of 100 bp
84132: 91442: contig of 7311 bp in length
91443: 91542: gap of 100 bp
91543: 94381: contig of 2839 bp in length
94382: 94481: gap of 100 bp
94482: 98181: contig of 3700 bp in length
98182: 98281: gap of 100 bp
98282: 102419: contig of 4138 bp in length
102420: 102519: gap of 100 bp
102520: 102520: contig of 4779 bp in length
102521: 107398: gap of 100 bp
107399: 110315: contig of 2917 bp in length
110316: 110416: gap of 100 bp
110417: 115743: contig of 5328 bp in length
115744: 115843: gap of 100 bp
115844: 119111: contig of 3268 bp in length
119112: 119211: gap of 100 bp
119212: 121219: contig of 2008 bp in length
121220: 121319: gap of 100 bp
121320: 123497: contig of 2178 bp in length
123498: 123597: gap of 100 bp
123598: 130224: contig of 6627 bp in length
130225: 130324: gap of 100 bp
130325: 135208: contig of 4884 bp in length
135209: 135308: gap of 100 bp
135309: 142309: contig of 7001 bp in length
142310: 142409: gap of 100 bp
142410: 144638: contig of 2229 bp in length
144639: 144738: gap of 100 bp
144739: 155510: contig of 10772 bp in length
155511: 155610: gap of 100 bp
155611: 160588: contig of 4978 bp in length
160589: 160688: gap of 100 bp
160689: 167138: contig of 6450 bp in length
167139: 167238: gap of 100 bp
167239: 172564: contig of 5326 bp in length
172565: 172664: gap of 100 bp
172665: 176139: contig of 3475 bp in length
176140: 176239: gap of 100 bp
176240: 178634: contig of 2395 bp in length
178635: 178734: gap of 100 bp
178735: 181546: contig of 2812 bp in length
181547: 181646: gap of 100 bp
181647: 183844: contig of 2198 bp in length
183845: 183944: gap of 100 bp
183945: 191043: contig of 7099 bp in length
191044: 191143: gap of 100 bp
191144: 195461: contig of 4318 bp in length
195462: 195561: gap of 100 bp
195562: 201981: contig of 6420 bp in length
201982: 202081: gap of 100 bp
202082: 204238: contig of 2157 bp in length
204239: 204338: gap of 100 bp
204339: 215768: contig of 11430 bp in length
215769: 215868: gap of 100 bp
215869: 218007: contig of 2139 bp in length
218008: 218107: gap of 100 bp
218108: 223568: contig of 5461 bp in length
223569: 223668: gap of 100 bp
223669: 226437: contig of 2769 bp in length
226438: 226537: gap of 100 bp
226538: 229262: contig of 2725 bp in length
229263: 229362: gap of 100 bp
229363: 229363: contig of 5869 bp in length
229364: 235231: gap of 100 bp
235232: 235331: contig of 2251 bp in length
235332: 237582: gap of 100 bp
237583: 237682: contig of 24848 bp in length
237684: 262530: gap of 100 bp
262531: 262630: contig of 7173 bp in length
262631: 269803: gap of 100 bp
269804: 272850: contig of 2947 bp in length

JOURNAL

Submitted (15-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 On Jun 18, 2000 this sequence version replaced gi:7622510.

COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 ----- Project Information -----
 Center project name: H NH0271N19
 ----- Summary Statistics -----
 Sequencing vector: M13; 70%
 Chemistry: Dye-terminator ET; 70% of reads
 Chemistry: Dye-terminator Big Dye; 30% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 180192 bases at least Q40
 Consensus quality: 181428 bases at least Q30
 Consensus quality: 182156 bases at least Q20
 Insert size: 182000; agarose-fp
 Insert size: 183849; sum-of-contigs
 Quality coverage: 5.14 in Q20 bases; agarose-fp
 Quality coverage: 4.97 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2056: contig of 2056 bp in length
 2057 2156: gap of unknown length
 2157 4432: contig of 2276 bp in length
 4433 4532: gap of unknown length
 4533 8368: contig of 3836 bp in length
 8369 8469: gap of unknown length
 8469 12467: contig of 3999 bp in length
 12468 12567: gap of unknown length
 12568 21302: contig of 8735 bp in length
 21303 31546: gap of unknown length
 31547 31647: gap of unknown length
 31647 47554: contig of 15908 bp in length
 47555 47655: gap of unknown length
 47655 77535: contig of 29880 bp in length
 77535 77635: gap of unknown length
 77635 105104: contig of 27470 bp in length
 105105 105204: gap of unknown length
 105205 139458: contig of 34254 bp in length
 139459 139558: gap of unknown length
 139559 184849: contig of 45291 bp in length.

FEATURES

source

1. .184849
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-271N19"
 1. .2056
 /note="assembly_name:Contig8"
 2157. .4432
 /note="assembly_name:Contig10"
 4533. .8368
 /note="assembly_name:Contig11"
 clone_end:SP6
 vector_side:left
 8469. .12467
 /note="assembly_name:Contig12"
 12568. .21302
 /note="assembly_name:Contig13"

misc_feature

21403. .31546
 /note="assembly_name:Contig14"
 31647. .47554
 /note="assembly_name:Contig15"

misc_feature

47655. .77534
 /note="assembly_name:Contig16"
 77635. .105104
 /note="assembly_name:Contig17"

misc_feature

clone_end:T7
 vector_side:left
 105205. .139458
 /note="assembly_name:Contig18"

misc_feature

139559. .184849
 /note="assembly_name:Contig19"

BASE COUNT 59891 a 32713 c 33432 g 57811 t 1002 others
 ORIGIN

Query Match 10.2%; Score 645.8; DB 2; Length 184849;
 Best Local Similarity 99.7%; Pred. No. 3.8e-132;
 Matches 647; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2548 TATTGGGCTTAGGTGTGATCATTTGCAATTTTGGATTAAATTTCTCCGAAGCTTTTAATCA 2607

Db 65412 TTTAGGGCTTAGGTGTGATCATTTGCAATTTTGGATTAAATTTCTCCGAAGCTTTTAATCA 65353

QY 2608 TTTTGGATGTGAGCCCTGCCAGTGAACCTCCATGGCTCAGTGAACAAATTTCTGCAATCC 2667

Db 65352 TGTGGATGTGAGCCCTGCCAGTGAACCTCCATGGCTCAGTGAACAAATTTCTGCAATCC 65293

QY 2668 TCACCTCTGGGAGTGTGAGTGCACAAAAGAGCCAAAGAGCTTCAGTGTGACACCTGCGAG 2727

Db 65292 TCACCTCTGGGAGTGTGAGTGCACAAAAGAGCCAAAGAGCTTCAGTGTGACACCTGCGAG 65233

QY 2728 AGAAAACTTTTATGGTTAGATGTCAACCAATTTGTAAGGCTCTGTGACTGTGACACAGCTGG 2787

Db 65232 AGAAAACTTTTATGGTTAGATGTCAACCAATTTGTAAGGCTCTGTGACTGTGACACAGCTGG 65173

QY 2788 ATCCCTCCCTGGGAGTGTCTGTAAATGTGAAGACAGGGCAGTGCATCTGCAAGCCCAATGT 2847

Db 65172 ATCCCTCCCTGGGAGTGTCTGTAAATGTGAAGACAGGGCAGTGCATCTGCAAGCCCAATGT 65113

QY 2848 TGAAGGGAGACAGTGCACATTAATTTGTTGAGGGAACTTTTACCTACGCGCAAAATAATTC 2907

Db 65112 TGAAGGGAGACAGTGCACATTAATTTGTTGAGGGAACTTTTACCTACGCGCAAAATAATTC 65053

QY 2908 TTTCTCTGTCTGCTGCTTGCACCTGTGAATAAGACTGGGACAAATAATTTGCTCTCTGCTGTG 2967

Db 65052 TTTCTCTGTCTGCTGCTTGCACCTGTGAATAAGACTGGGACAAATAATTTGCTCTCTGCTGTG 64993

QY 2968 TAACAAATCAACAGGACCAATGTCTTGCACAAATTAAGGGGTAAACAGGTCTTTCGCTGTAATCA 3027

Db 64992 TAACAAATCAACAGGACCAATGTCTTGCACAAATTAAGGGGTAAACAGGTCTTTCGCTGTAATCA 64933

QY 3028 GTGTGAGCTTCAAGGTACAATTTGACCAATTTTCAACACTCCAGAGTGTGCA 3087

Db 64932 GTGTGAGCTTCAAGGTACAATTTGACCAATTTTCAACACTCCAGAGTGTGCA 64873

QY 3088 GTGTGATTTCTTGGGGACATTTACCTGGGACCAATTTGTGACCCCAATCAGTGGCAGTGCCT 3147

Db 64872 GTGTGATTTCTTGGGGACATTTACCTGGGACCAATTTGTGACCCCAATCAGTGGCAGTGCCT 64813

QY 3148 GTGTGTGCTTAATCTGTAAGGAAGAGGTGTAATCAGTGTCAACCCAGGT 3196

Db 64812 GTGTGTGCTTAATCTGTAAGGAAGAGGTGTAATCAGTGTCAACCCAGGT 64764

RESULT 12

LOCUS

AL161641 290708 bp DNA linear HTG-23-OCT-2002
 DEFINITION Homo sapiens chromosome 1 clone RP5-865023 map p35.3-36.12, ***
 SEQUENCING IN PROGRESS ***, 56 unordered pieces.
 ACCESSION AL161641
 VERSION AL161641.20 GI:23337111

Db 526 AGACCAACCAATGTGATGTTCAACAGGAGTGTAGTGTCACTTACAGACCAAAACAG 585
Qy 3822 GGGTCCGAGAGGAACTTGACTTTAAGTTATATATCTTATGCTCAGACTCTGTGA 3881
Db 586 GGGTCCGAGAGGAACTTGACTTTAAGTTATATATCTTATGCTCAGACTCTGTGA 645
Qy 3882 CACTTACCTGGACAACTCTCAATCAATCTGCTCCATAGAGAAATATATTTGCTCT 3941
Db 646 CACTTACCTGGACAACTCTCAATCAATCTGCTCCATAGAGAAATATATTTGCTCT 705
Qy 3942 GTGCCCTTTGGCTGGTCTGAGCCATGTTCTTCCAGAAAGTCAATGAACTCTCAGCTA 4001
Db 706 GTGCCCTTTGGCTGGTCTGAGCCATGTTCTTCCAGAAAGTCAATGAACTCTCAGCTA 765
Qy 4002 CCATCTGCAATCTGCTTCCATTTGCTGCAAGTACATTTTCTGTACAGGGTGTACTAGCG 4061
Db 766 CCATCTGCAATCTGCTTCCATTTGCTGCAAGTACATTTTCTGTACAGGGTGTACTAGCG 825
Qy 4062 GGGGCTGTTTACACAGCTTGGCCATTTACAGTCAACACAGCCAGCCCTCCCAAGAC 4121
Db 826 GGGGCTGTTTACACAGCTTGGCCATTTACAGTCAACACAGCCAGCCCTCCCAAGAC 885
Qy 4122 TAAGTCCACCTAAGATGAGCAAAATCAGTTTCTACAGAACTTCAATGAGTGTCTCCAC 4181
Db 886 TAAGTCCACCTAAGATGAGCAAAATCAGTTTCTACAGAACTTCAATGAGTGTCTCCAC 945
Qy 4182 CAGCGGAACCTAAGTCAATATTAAG 4209
Db 946 CAGCGGAACCTAAGTCAATATTAAG 973

RESULT 14
AX347325/c
LOCUS AX347325 2749 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 2396 from Patent WO0200928.
ACCESSION AX347325
VERSION AX347325.1 GI:18495213

KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 2396 03-JAN-2002;
Epigenomics AG (DE)

FEATURES
Location/Qualifiers
source
1..2749
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 842 a 25 c 536 g 1346 t
ORIGIN

Query Match 7.5%; Score 476.4; DB 6; Length 2749;
Best Local Similarity 80.6%; Pred. No. 1.3e-94;
Matches 569; Conservative 0; Mismatches 136; Indels 1; Gaps 1;

Qy 183 ATACGACGAGTACTCATGCTTCCGCAATGCTAAGAACGCTGTTGGTATTACCTTACTC 242
Db 1228 ATACGACGAGTACTCATGCTTCCGCAATGCTAAGAACGCTGTTGGTATTACCTTACTC 1169
Qy 243 TGAGAACGTGTCTGAGTTTCCAGAAATGGAGTATCGCAACATCATTAAAGTACCCCTG 302
Db 1168 TAAAGACGTGTCTGAGTTTCCAGAAATGGAGTATCGCAACATCATTAAAGTACCCCTG 1109
Qy 303 CTTCAGAAATGCTGGGAGTGGGCTGATTTATTTATTAAGAAATGCTTTATCA 362
Db 1108 CTTCAGAAATGCTGGGAGTGGGCTGATTTATTTATTAAGAAATGCTTTATCA 1049
Qy 363 CGAGGAGAAATGC-TTTTGTGTAACATGAATGCGCCAGTCTTTCATTGGGCTCTGGCTC 421

Db 1048 AAAAAAATACTCTTTTATAAACATAAATACCAATCTCTTCAATAAATCTTAACCTTC 989
Qy 422 TTGTTTCAGGTCATTGAATCTTTCCTTATTTGCTTCAATATCTTCAATATCTTCACTGAG 481
Db 988 TTAATTCAAATCAATAAATATTAATCTTTACCTTATTTCTTCAATATCTTCACTTAA 929
Qy 482 TCAGAGGTCTTTTCCAGGCTGGAGAACGCTGGAGGCTTTCAAGAAATGTTCCATCGTG 541
Db 928 TCAGAAATCTTTTCCCAAACTAAATAACGCTTCAAAAAATTTCCATCGTA 869
Qy 542 CCAACCCAAAGCAGTATGTGGACTCCAGAGCCGAGGACCTTTTGTGTCAGAGCTCTGCTCT 601
Db 868 CCAACCCAAAGCAGTATGTGGACTCCAGAGCCGAGGACCTTTTGTGTCAGAGCTCTGCTCT 809
Qy 602 GCTGAAAGTATTCAGTTCTGTACCCAGCGGTTTGTATTCAGGATTTGCCCATACAGATCT 661
Db 808 ACTAAATAATTTCAATCTATACCAACGATTTTATTTCAAAATTTACCATACAAATCT 749
Qy 662 TCACACCTTACCTACCTACCTGCTTTTCTCAGAGGCTCTCAGTGTGATCAGACGAGAC 721
Db 748 TCACACCTTACCTACCTACCTGCTTTTCTCAGAGGCTCTCAGTGTGATCAGACGAGAC 689
Qy 722 AAGATGATCTGCACTCTTAAGCCCATAGCAATTTCTCAAGTCTTTATTTTGGAAATCAC 781
Db 688 AAAAAATCTATACCTTACCTTACCGCCCAACATTTCTACAAATTTTATTTTAAAAATCAC 629
Qy 782 AAGAGCTGCTTTTCTCTCTCTCTTCTCCTTCTCCTTCTCCTTCTCCTTCTTACCTTACTGTA 841
Db 628 AAAAACTACTTTTCTCTCTCTCTTCTCCTTCTCCTTCTCCTTCTCCTTCTTACTTACTATA 569
Qy 842 TGGCTGAACTGAGCAACAGGCTGATGTTGTTTATATAGAAAAGA 887
Db 568 TAACTAAACCTTAAACCAACAAATATATATATAATATAATATAATAA 523

RESULT 15
AX347324
LOCUS AX347324 2749 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 2395 from Patent WO0200928.
ACCESSION AX347324
VERSION AX347324.1 GI:18495212

KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.

REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 2395 03-JAN-2002;
Epigenomics AG (DE)

FEATURES
Location/Qualifiers
source
1..2749
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 810 a 25 c 561 g 1353 t
ORIGIN

Query Match 6.9%; Score 437; DB 6; Length 2749;
Best Local Similarity 76.5%; Pred. No. 6.9e-86;
Matches 547; Conservative 1; Mismatches 166; Indels 1; Gaps 1;

Qy 182 GATACACGAGTACTCATGCTTCCGCAATGCTAAGAACGCTGTTGGTATTACCTTACT 241
Db 1521 GATATAGTATGTTTATTTATTTTCTTATTTAGAACGCTGTTGGTATTATTTAT 1580
Qy 242 CTGAGAACGTTCTGAGTTTCCAGAAATGGAGTATCGCAACATCATTAAAGTACCCCT 301
Db 1581 TTGAGAACGTTCTGAGTTTCCAGAAATGGAGTATCGTAATATTTTAAAGTATTTT 1640
Qy 302 GCTTCAAGATTTGCTGGCAAGTGGGCTGATTTATTTATTTAGAAATGCTTTATC 361


```

Db      1641 GTTTAAAGTATTCTGGTAAGTGGCGTGGTTTGGATTATTATTAGAAATGTTTATT 1700
Qy      362 AGGAGGAGATG-CTTTTGTAAACATGATGCGGAGTCTCTTCATGGGCTCTGCTT 420
Db      1701 AGGAGGAGATGTTTTTTTGTAAATATGAATGTGTAGTTTTTTTATGGGTTTGGTTT 1760
Qy      421 CTTGTTTCAGCTCATTGAAATGTTGATCTTTGCTATTTTGGCTTCAATATCCTTGACTGA 480
Db      1761 TTTGTTTAGTATTGAATGTTGATTTTGTATTATTTTGTATTAAATTTTGTATTGA 1820
Qy      481 GTACAGAGTCTTTTCCAGAGGCTGAGAACTGGGAGCTTTCAGAAAGTTTCCATCGT 540
Db      1821 GTTACGAGGTTTTTTTTTAAAGTTGGAGAACGTGGAGTTTTTAAGAAAGTTTTATCGT 1880
Qy      541 GCCAACCCACAGCAGTATGCGACTCCAGACCGAACACTTTTTGTGCACAGCTCTGCTGC 600
Db      1881 GTTAATTTAAGTAGTAGTGAGTTTTTGTATCGAAGTATTTTTGTATAGTTTTGTGT 1940
Qy      601 TGCTGAAAGTATTTCAGTTCTGTACCCAGCGGTTTTGTATTTCAGGATTTGCCCATACAGTC 660
Db      1941 TGTGAAAGTATTTAGTTTTGTATTAGCGGTTTTGTATTAGGATTTTATATAGATT 2000
Qy      661 TTCACACCTTACCTACACTGCCCTTTTCTCAGAGGCTCAGTAGCTGCATCACACCAGA 720
Db      2061 1 TTTATATTTTATTATATATGTTTTTTTTTTTAGTAGGTTTTTAGTAGTTGTATTATAGA 2060
Qy      721 CAAGAATGATCTGCATCCTTAACGCCCATAGCAATTCGCAAGTTTTTATTTTGGAAATCA 780
Db      2061 TAAGAATGATTTGTATTTTAAAGTTTATATAGTAAATTTGTAAAGTTTTTATTTTGGAAATTA 2120
Qy      781 CAAGAGCTGTTTTCTTCTCTCTCTCCAAAGCTGATGGCATCATTTACCTTAGCTGT 840
Db      2121 TAAGAGTGTTTTTTTTTTTTTTTTTTTTTTTTAAAGTTGATGGTATTATTATTAGTTGT 2180
Qy      841 ATGGCTGAACCTGAGCAACAAGGTGTAATGTGTATTATAGAAAGACRGTAGAT 895
Db      2181 ATGGTTGAATTTAGTAATAAGGTGTAATGTAAGTAGTAGTAGTAGTATAGTT 2235

```

Search completed: January 1, 2004, 11:23:45
 Job time : 14855 secs